

An avatar-based system for identifying individuals likely to develop dementia

Bahman Mirheidari¹, Daniel Blackburn², Kirsty Harkness³, Traci Walker⁴, Annalena Venneri^{2,5}, Markus Reuber⁶, and Heidi Christensen¹

¹Department of Computer Science, University of Sheffield, Sheffield, UK

²Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK

³Department of Neurology, Royal Hallamshire Hospital, Sheffield, UK

⁴Department of Human Communication Sciences, University of Sheffield, Sheffield, UK

⁵IRCCS Fondazione Ospedale San Camillo, Venice, Italy

⁶Academic Neurology Unit, University of Sheffield, Royal Hallamshire Hospital, Sheffield, UK

{bmirheidari2,d.blackburn,traci.walker,a.venneri,m.reuber,
heidi.christensen}@sheffield.ac.uk, kirsty.harkness@sth.nhs.uk

Abstract

This paper presents work on developing an automatic dementia screening test based on patients' ability to interact and communicate - a highly cognitively demanding process where early signs of dementia can often be detected. Such a test would help general practitioners, with no specialist knowledge, make better diagnostic decisions as current tests lack specificity and sensitivity. We investigate the feasibility of basing the test on conversations between a 'talking head' (avatar) and a patient and we present a system for analysing such conversations for signs of dementia in the patient's speech and language. Previously we proposed a semi-automatic system that transcribed conversations between patients and neurologists and extracted conversation analysis style features in order to differentiate between patients with progressive neurodegenerative dementia (ND) and functional memory disorders (FMD). Determining who talks when in the conversations was performed manually. In this study, we investigate a fully automatic system including speaker diarisation, and the use of additional acoustic and lexical features. Initial results from a pilot study are presented which shows that the avatar conversations can successfully classify ND/FMD with around 91% accuracy, which is in line with previous results for conversations that were led by a neurologist.

Index Terms: clinical applications of speech technology, pathological speech

1. Introduction

Dementia is a disorder of the brain, caused by a number of different pathological processes including Alzheimer's disease (AD). In 2015 it was estimated that 46.8 million people around the world had dementia, and this number is expected to double in 20 years. AD and most other forms of dementia predominantly affect the neuropsychological domains of learning and memory but speech and language are also affected. Much work on language decline in AD focuses on performance in simple quantitative tests over time. Less research has concentrated on the conversational contributions of patients with dementia, despite the fact that the diagnosis is usually suspected and made through the interaction of the patient and a clinician.

A routine part of any presentation of a new medical problem to a doctor is the *history taking*, where an examiner (the doctor) asks a number of specific questions of a patient, observes and interprets the responses. This whole process is usually carried out in a semi-structured manner and analysed informally in a qual-

itative fashion (leaving the clinician with an *impression* that a particular patient is or is not likely to be developing a form of dementia). At present, there is no validated automatic tool capable of aiding or quantifying the spoken communication capacity of the patient. In patients, in the early stages of a dementing illness, the detection of subtle diagnostic clues requires considerable expertise, likely to be lacking in the non-specialist medical settings in which patients initially present with their cognitive concerns. This means that conversational clues to developing dementia may be missed and that screening errors are likely to be made. Such errors will result in incorrect referral decisions, so that patients who are highly functioning (e.g., present as very articulate) are not referred to specialist services for the investigation of possible dementia (false negatives). Conversely, other patients, who are worried about having a serious memory problem because of ordinary memory failings but who are not in the process of developing dementia, are referred to specialist memory clinics (false positives), adding to the strain on limited expert services for patients with dementia. Simple pen-and-paper tests, used routinely at the moment to identify patients at risk of developing dementia in primary care settings, have suboptimal sensitivity and specificity. Therefore, it is highly desirable to improve the diagnostic process by combining established cognitive screening tests with an automatic process capable of identifying individuals at risk of developing dementia.

In a recent study, Elsey *et al.* applied Conversation Analysis (CA) to neurologist-patient interactions recorded in memory clinics and found that profiles based on interactional contributions of patients' (and that of accompanying others) could be used to distinguish between patients with memory problems due to an emerging neurodegenerative disorder (ND) and those with Functional Memory Disorder (FMD) (non-progressive memory complaints not associated with a neurological condition) [1, 2]. The study showed promising results in terms of diagnostic accuracy, but relied on manual CA for the discovery of the interaction patterns in the conversation. The CA process involves a number of steps including audio recording, manual transcription and completion of a qualitative analysis by a trained expert. It is thus prohibitively expensive and time consuming, and not feasible for large-scale use. Examining the same recordings and transcripts analysed in the CA study, we were able to demonstrate that the process can be automated [3, 4]. Using speech recognition, manual diarisation and some speech understanding, we obtained an FMD/ND classification accuracy of over 90% when applying feature selection.

The present paper builds on this work and aims to advance the process of fully automating a diagnostic procedure based on the combination of cognitive screening tests and the analysis of interaction. To this end we wanted to investigate whether it is possible to create a conversation-based diagnostic test using an animated head on a computer screen (an *avatar*) instead of relying on a clinician to lead the conversation. The automation of the recording of a patient's communication behaviour would greatly increase the utility of the tool and allow it to be used in non-expert clinical settings or even in a patient's own home. Access to such a low-cost, easy-to-use test would allow the patient selection procedure for specialist investigation or treatment to be improved and limited specialist resources to be deployed more effectively. An automated testing procedure would also facilitate repeat testing, for instance to identify change over time or to capture variability during the course of the day.

The paper is structured as follows. Section 2 gives an overview of previous work on automatic dementia detection and presents our system, Section 3 describes the experimental setup including the parallel neurologist-patient and avatar-patient corpora. Finally, results and conclusions are presented in Sections 4 and 5 respectively.

2. Automatic detection of dementia

Whilst the automatic analysis of interaction is quite a new field of study [5], which has rarely been applied to the differential diagnosis of memory problems, a significant amount of work has been carried out using machine learning techniques to identify signs of dementia in patient's speech and language. Lopez-de-Ipina *et al.* [6, 7] investigated a number of acoustic features (duration, time domain and frequency domain), and the fractal dimension from the multilingual AZTIAHO database. They obtained binary classification accuracy rate of around 90% to 95% between the patients with AD and healthy adults.

Toth *et al.* [8] have found other acoustic and lexical features (e.g. length of utterance and pauses) useful in identifying patients with mild cognitive impairment (MCI). In their recent work, Gosztolya *et al.* [9] expanded the initial feature set to include descriptors (silence and filled pauses, breathing noises, laughter and coughs). Applying a number of different feature selection algorithms, they have tried to identify the most informative features for classification. Jerrold *et al.* [10] combined half of their Automatic Speech Recognition (ASR) outputs with half of their human transcriptions to extract acoustic and lexical features. The binary classification accuracy between AD and healthy controls was around 88%. Likewise, Thomas *et al.* [11] saw a drop in accuracy when attempting to distinguish between more than one type of dementia. [12] used the manual transcripts of the conversations between the interviewers and the patients (14 with MCI and 27 with cognitively intact problems) to extract a number of lexical features. The best classification obtained by the SVM classifier with 88% accuracy rate.

Satt *et al.* [13] carried out a study on subjects with MCI, AD and healthy adults, asking them to complete tasks like verbally describing a picture while looking at it. In a study on data collected from the interdisciplinary longitudinal study on adult development and ageing (ILSE, a German collection of 1000 participants' spontaneous speech in their middle adulthood and later life), Weiner *et al.* [14] extracted a number of acoustic and linguistic features (e.g. silence duration, word and phoneme rates) to train a classifier distinguishing between three categories: AD, ageing-associated cognitive decline (AACD) and healthy adults. To extract the lexical features, they used manual

transcriptions and for the acoustic features they applied a voice activity detection (VAD) technique. Using a linear discriminant analysis (LDA) classifier, they obtained 85.7% classification accuracy for the three participant groups. Whilst differentiating between healthy participants and those with AD was successful, the classifier was not capable of categorising the healthy group from the AACD patients. [15, 16] have used the DementiaBank corpus (containing speech of patients with AD, vascular dementia, MCI and healthy controls describing the 'Cookie Theft' picture) to predict changes in patients' Mini Mental State Examination (MMSE) scores over time. The researchers extracted a wide range of features (477 lexico-syntactic, acoustic, and semantic) and selected the 40 most informative, reporting an accuracy of over 92% in terms of the distinction of AD patients from healthy controls. However, in their very latest study [17], using state-of-the-art ASR (instead of manual transcripts) with 38.2% word error rate (WER) and extracting only the lexical features, the accuracy of the classifier dropped significantly. Working on the same dataset, [18] achieved a 94% classification accuracy using acoustic-only features.

In brief, recent research has demonstrated that automatic audio and speech technology may provide diagnostic markers that can aid the classification between e.g., HC and people with AD or MCI. However, most studies have focused on providing a supplementary, automatic method based on existing test procedures currently used in clinical settings like picture description. In addition, many research studies have used manual transcription, thereby side-stepping the known challenges associated with the fully automated analysis of spontaneous speech.

In contrast, we propose a fully automated system to detect early interactional signs of dementia (Figure 1). First, an audio file containing a recording of the conversation is passed to a diarisation tool to identify the speech portions of the input audio stream, as well as the speaker ID of each speech segment. This information is then passed to an ASR system. The ASR is given both the input audio file and the output produced by the diarisation tool to generate a string of words spoken by each speaker. Next, the output of the diarisation tool and the ASR are given to the feature extraction unit to extract a number of features. Some features may rely on techniques such as signal-, text- and natural language processing as well as spoken language understanding. Finally, the extracted features are sent to a machine learning classifier to decide which category the whole conversation belongs to; in this study the two groups are patients diagnosed with neurodegenerative dementia (ND) and patients with Functional Memory Disorder (FMD) [19]. The diarisation, ASR and classification are further described in Sections 3.2 and 4 respectively.

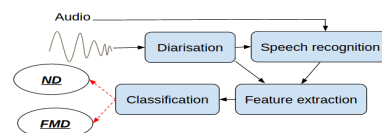


Figure 1: Automatic dementia detection system.

3. Experimental setup

3.1. Data

This study has used data from two parallel corpora both of which were recorded at the Royal Hallamshire Hospital's memory clinic in Sheffield, UK.

3.1.1. Neurologist-patient conversations

The database of neurologist-patient conversations consists of 30 (15 FMD and 15 ND) audio recordings and associated manual CA annotations. For this study, we had access to an additional dataset of 24 conversations for the training of the speech recogniser. Patients were encouraged to bring an accompanying person with them to the memory clinic, so many conversations included three participants ‘Neu’ (neurologist), ‘Pat’ (patient) and ‘Aps’ (accompanying person(s)). The neurologists were instructed to attempt to follow a predefined set of questions constructed with the intent to give patients an opportunity to reveal the typical signs of interactional impairments in the conversation. Several categories of questions were used including closed and open-ended questions, compound questions and questions related to memory concerns. Further details about the questions can be found in [1] and [4]. As the data was recorded in an ordinary clinical setting and not recorded with the aim of applying speech recognition, little effort was made to reduce background noise and acoustic interference, and for many of the recordings the microphone placement was relatively *ad hoc* (often being placed closer to the neurologist than the patient). In addition, the speech itself was very challenging with a high percentage of overlapping speech segments on occasion even professional transcribers has not been able to transcribe the material.

3.1.2. Avatar-patient conversations

For the current study, aimed at investigating to what degree it is feasible to use an avatar front-end to elicit conversational diagnostic features, a prototype avatar was developed (Figure 2).



Figure 2: Prototype avatar (using <https://www.botlibre.com>).

Questions similar in style to the ones used in the neurologist-patient conversations were put together in collaboration with a clinical linguist. For this initial study, these questions were recorded by an American male but for future work we plan to replace this with a synthetic voice. Participants were given instructions about how to use the software on a laptop; to minimise confusion, most of the keys on the keyboard were covered leaving only labelled ‘next’ and ‘play’ keys visible. 24 participants took part and out of these, a total of 12 recordings of patients interacting with the avatar were further analysed (6 ND, 6 FMD), as we excluded 4 with depressive pseudo dementia, 6 with MCI and 2 in whom the diagnosis was not clear. Audio was recorded using the laptop’s built-in microphone. We also recorded two video streams from the built-in webcam and from a webcam positioned to the side of the participants. In the current study, only the audio has been used. As with the neurologist-led conversations, the patient were allowed to bring an accompanying person, if they wanted to.

3.2. Feature extraction

CA inspired features Similar to the previous studies ([3, 4]), a total number of 20 features were extracted from the outputs of the diarisation and the ASR modules, to replicate the quali-

tative features proposed by Elsey *et al.* [1] using CA as closely as possible. Features were extracted individually for each of the conversation participants and named accordingly using prefixes: ‘Neu’, ‘Pat’ and ‘Aps’ (see the first row of Table 1). The linguistic features were collected using Bag-of-Words (BoW) [20] and the NLTK python library ([21]). For the *conceptual* features, a simple approach of searching for predefined keywords were used. In addition to the CA inspired features, two more groups of features were extracted from the patients’ turns only: lexical (part of speech) and more acoustic features.

Lexical (Part of Speech) features Penn Treebank part of speech tags ([22]) were assigned to the words uttered by the patients in the conversations. The number of the Penn Treebank tags are originally 36, however, the similar tags (e.g. different types of verbs) were joined together to make more general tags. The tags were gathered under 12 different groups (Table 1; second row).

Acoustic features Using the well known ‘Praat vocal toolkit’ ([23]), the total number of 12 acoustic features were extracted from the audio recordings of the patients in the conversations. We were interested in features that are usually marked in formal CA transcripts, including the prosodic features (duration, pitch and intonation), creakiness and breathiness (H1-H2 [24], H1-A1 [25], H1-A2, H1-A3) and vocal stability (jitter, shimmer, harmonics-to-noise and noise-to-harmonics ratios).

4. Results

Table 2 presents the results of training various ASR models on the neurologist-patient (HUM) and avatar-patient (AVA) conversations using the Kaldi toolkit [26] (acoustic model: SAT trained HMM-GMM; language model: based on training dataset using Kneser smoothing). In both cases, all conversations available are used for training and models are tested and adapted using only the respective test sets contained the conversations with appropriate diagnoses as described in Section 3.1.

As described above, this data is very challenging because of recording conditions, the spontaneity of the conversations and the potential speech and language deficiencies of the patient. As a result, the baseline WERs for the conversations are high at 55.7% (HUM) and 77.0% (AVA) (first two rows in Table 2¹). The very high AVA result is likely to be caused by insufficient data; in fact, decoding the AVA set on the HUM baseline model (row 3 in Table 2) by itself improves the result to 65.0%. This is further improved upon (rows 4 and 5) by applying Maximum a posteriori (MAP) adaptation [27] and by simply combining the two datasets (i.e. with more data for training) which gives the best result at 46.2%. All models were trained and tested using the leave-one-out principle. We applied diarisation to the two datasets using the SHoUT ([28]) toolkit and obtained diarisation error rates of 30.4% and 29.6% respectively. Although seemingly high, when looking at the corresponding rate of words attributed to the wrong speaker, the percentages are only 5.7% for HUM and 16.8% for AVA.

Table 3 shows the classification results for the HUM and AVA conversations for the individual groups of features, for all features and when applying the top 10 features (by Recursive Feature Elimination (RFE) method [29]) as derived from the HUM data. Depending on the degree of automatising, different types of features are more useful, e.g., when using the manual

¹The HUM result is significantly higher than the results presented in [3] due to problems with scoring and diagnosis uncertainty which has since been rectified.

Table 1: List of extracted features. Prefixes: ‘Neu’ (neurologist), ‘Pat’ (patient), ‘Aps’ (accompanying person(s)).

Category	Feature
CA inspired	number of turns (APsNoOfTurns , PatNoOfTurns , NeuNoOfTurns); average length of turn (APsAVTurnLength , PatAVTurnLength , NeuAVTurnLength); number of unique words in a turn (APsAVUniqueWords , PatAVUniqueWords , NeuAVUniqueWords); patient answers “me” for question “who’s most concerned” (PatMeForWhoConcerns); patient recalls memory failure features (PatFailureExampleEmptyWords , PatFailureExampleAVPauses , PatFailureExampleAllTime); patient replies ‘dunno for the expectation question (PatDunnoForExpectations); average number of filler, empty, unique and low-frequency words (PatAVFillers), PatAVEmptyWords , PatAVUniqueWords PatAVAllWords); average number of repeated questions (AVNoOfRepeatedQuestions); average number of topics discussed (AVNoOfTopics)
Lexical (Part of speech)	average number of verbs, nouns, adjectives, adverbs, pronouns, wh_words(e.g. who), determiner, conjunctions, cardinals, existential(e.g., there is), prepositions etc(PatAvgVerb , PatAvgNoun , PatAvgAdjective , PatAvgAdverb , PatAvgPronoun , PatAvgWh_word , PatAvgDeterminer , PatAvgConjunction , PatAvgCardinal , PatAvgExistential , PatAvgPreposition , PatAvgOtherPOS)
Acoustic	average overall intonation, pitch, duration and silence(PatAvgIntonation , PatAvgPitch , PatAvgDuration PatAvgSil); difference between the first harmonic and the harmonic close to the first, second and third formants(PatAvgH1-A1 , PatAvgH1-A2 , PatAvgH1-A3); difference between the two first harmonics (PatAvgH1-H2); local jitter and shimmer(PatAvgGitterLocal , PatAvgShimmerLocal); harmonics-to-noise and noise-to-harmonics ratios(PatAvgMeanHNR , PatAvgMeanNHR)

Table 2: Speech recognition results.

System	Train	Test	WER
Baseline_HUM	HUM	HUM	55.7%
Baseline_AVA	AVA	AVA	77.0%
Cross domain	HUM	AVA	65.0%
MAP adaptation	Map on HUM	AVA	58.7%
Combining data	HUM+AVA	AVA	46.2%

transcripts for AVA (2nd row in Table 3), we get the highest results using the lexical features but their usefulness drops as the ASR transcripts (with associated recognition errors) are introduced. The best result for the AVA conversations (90.9%) is obtained when using all features, which is in line (or even above) what we achieved with for the HUM conversations. It is clear that replacing the neurologist with an avatar can still lead to conversations in which signs of dementia is present.

Some differences were observed between the AVA and HUM data. Figure 3 shows a plot of the average length of the turns in the conversations for ND and FMD groups respectively. In both datasets the ND group’s patients had shorter turns however, overall the AVA conversations had much longer turns, which is likely to be because this initial avatar provides no feedback to the patients in the form of nods, clarifying questions or back-channel noises to steer the conversation.

The subjective survey at the end of the experiment showed that the overall participants’ feedback about the avatar was very positive with a high level of satisfaction - some patients even indicated that they would prefer an automated test like this as it would feel less intimidating.

5. Conclusions

We have proposed a fully automatic system for detecting dementia in conversations between patients and neurologists, and demonstrated the feasibility of replacing the neurologists with an avatar. This is an essential first step for developing a low-cost tool for the early diagnosis of dementia as well as for post-diagnosis tracking of patients’ deterioration. The low cost of the

Table 3: Classification accuracy; ‘_man’:using gold-standard transcript instead of ASR-produced transcripts; ‘CA’: CA-style features; ‘AC’:acoustic features; ‘LX’:lexical features; ‘T10’:top 10 informative features.

Train/Test	CA	AC	LX	ALL	T10
HUM_man/ HUM_man	96.7%	83.3%	66.7%	76.7%	100%
HUM/ HUM	76.7%	60.0%	50.0%	76.7%	90.0%
AVA_man+ HUM_man/AVA_man	58.3%	66.7%	83.3%	66.7%	75.0%
AVA_man+ HUM_man/AVA	72.7%	63.6%	63.6%	81.8%	72.7%
AVA+ HUM_man/AVA	63.6%	54.5%	63.6%	90.9%	72.7%

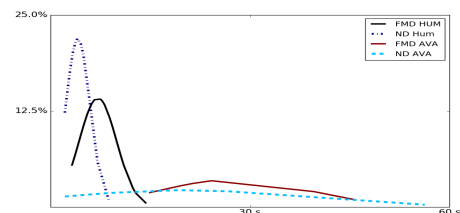


Figure 3: Histogram of the average turn length.

potential tool to stratify patients with memory complaints is important as prevalence of cognitive impairment is increasing with an aging society and funding constraints. In future work we will improve the avatar front-end to make it appear more responsive to the style and content of the patients’ answers and trial the tool with a larger cohort and in multiple settings including people’s homes.

6. References

- [1] C. Elsey, P. Drew, D. Jones, D. Blackburn, S. Wakefield, K. Harkness, A. Venneri, and M. Reuber, "Towards diagnostic conversational profiles of patients presenting with dementia or functional memory disorders to memory clinics," *Patient Education and Counseling*, vol. 98, pp. 1071–1077, 2015.
- [2] D. Jones, P. Drew, C. Elsey, D. Blackburn, S. Wakefield, K. Harkness, and M. Reuber, "Conversational assessment in memory clinic encounters: interactional profiling for differentiating dementia from functional memory disorders," *Aging & Mental Health*, vol. 7863, pp. 1–10, 2015.
- [3] B. Mirheidari, D. Blackburn, M. Reuber, T. Walker, and H. Christensen, "Diagnosing people with dementia using automatic conversation analysis," in *Proceedings of Interspeech*. ISCA, 2016, pp. 1220–1224.
- [4] B. Mirheidari, D. Blackburn, K. Harkness, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Towards the automation of diagnostic conversation analysis in patients with memory complaints," *Journal of Alzheimer's Disease*, 2017.
- [5] R. J. Moore, "Automated Transcription and Conversation Analysis," *Research on Language and Social Interaction*, vol. 48, no. 3, pp. 253–270, 2015.
- [6] K. Lopez de Ipina, J.-B. Alonso, C. M. Travieso, J. Sole-Casals, H. Egiraun, M. Faundez-Zanuy, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martinez-Lage, and U. Martinez de Lizardui, "On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis," *Sensors*, vol. 13, pp. 6730–45, 2013.
- [7] K. Lopez de Ipina, J. Sole-Casals, H. Egiraun, J. B. Alonsod, C. M. Travieso, A. Ezeiza, N. Barroso, 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," *Computer Speech and Language*, vol. 30, pp. 43–60, 2015.
- [8] L. Toth, G. Gosztolya, V. Vincze, I. Hoffmann, G. Szatloczki, E. Biro, F. Zsura, M. Pakaski, and J. Kalman, "Automatic detection of mild cognitive impairment from spontaneous speech using ASR," *Interspeech*, 2015.
- [9] G. Gosztolya, L. Tóth, T. Grósz, V. Vincze, I. Hoffmann, G. Szatloczki, M. Pákaski, and J. Kálmán, "Detecting Mild Cognitive Impairment from Spontaneous Speech by Correlation-Based Phonetic Feature Selection," *Interspeech 2016*, pp. 107–111, 2016.
- [10] 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," *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 27–37, 2014.
- [11] C. Thomas, V. Keselj, Cercone, K. Rockwood, and E. Asp, "Automatic detection and rating of dementia of Alzheimer type through lexical analysis of spontaneous speech," *Proceedings of the IEEE International Conference on Mechatronics & Automation*, pp. 1569–1574, 2005.
- [12] M. Asgari, J. Kaye, and H. Dodge, "Predicting mild cognitive impairment from spontaneous spoken utterances," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 2017.
- [13] A. Satt, A. Sorin, O. Toledo-Ronen, O. Barkan, I. Kompatsiaris, A. Kokonozi, and M. Tsolaki, "Evaluation of speech-based protocol for detection of early-stage dementia," *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, pp. 1692–1696, 2013.
- [14] J. Weiner, C. Herff, and T. Schultz, "Speech-Based Detection of Alzheimer's Disease in Conversational German," *Interspeech 2016*, pp. 1938–1942, 2016.
- [15] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic Features Identify Alzheimer's Disease in Narrative Speech," *Journal of Alzheimer's Disease*, vol. 49, pp. 407–22, 2015.
- [16] 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.
- [17] L. Zhou, K. C. Fraser, and F. Rudzicz, "Speech recognition in Alzheimer's disease and in its assessment," *Proceedings of the 17th Annual Meeting of the International Speech Communication Association (Interspeech)*, pp. 1948–1952, 2016.
- [18] S. Al-Hameed, M. Benaissa, and H. Christensen, "Simple and robust audio-based detection of biomarkers for alzheimer's disease," *SLPAT 2016: 7th Workshop on Speech and Language Processing for Assistive Technologies*, pp. 32–36, 2016.
- [19] K. Schmidtke, S. Pohlmann, and B. Metternich, "The syndrome of functional memory disorder: definition, etiology, and natural course," *Am J Geriatr Psychiatry*, vol. 16, pp. 981–8, 2008.
- [20] G. Salton, *Introduction to modern information retrieval*. New York, London, McGraw-Hill, 1983.
- [21] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media Inc, 2009.
- [22] A. Taylor, M. Marcus, and B. Santorini, "The penn treebank: an overview," in *Treebanks*. Springer, 2003, pp. 5–22.
- [23] P. P. G. Boersma *et al.*, "Praat, a system for doing phonetics by computer," *Glott international*, vol. 5, 2002.
- [24] M. Gordon and P. Ladefoged, "Phonation types: a cross-linguistic overview," *Journal of Phonetics*, vol. 29, no. 4, pp. 383–406, 2001.
- [25] S. u. D. Khan, K. Becker, and L. Zimman, "The acoustics of perceived creaky voice in american english," *The Journal of the Acoustical Society of America*, vol. 138, no. 3, pp. 1809–1809, 2015.
- [26] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, Dec. 2011, iEEE Catalog No.: CFP11SRW-USB.
- [27] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," *IEEE transactions on speech and audio processing*, vol. 2, no. 2, pp. 291–298, 1994.
- [28] M. Huijbregts, "Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled," Ph.D. dissertation, University of Twente, The Netherlands, 2008.
- [29] F. Pedregosa and G. Varoquaux, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.