Automated Speech-based Screening for Alzheimer's Disease in a Care Service Scenario

Johannes Tröger, Nicklas Linz and Jan Alexandersson
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
<f>.<l>@dfki.de

Alexandra König and Philippe Robert
Memory Center, CoBTeK - IA CHU Université Côte d'Azur, France
Nice, France
akonig03@gmail.com,
phil.robert15@orange.fr

ABSTRACT
This paper describes a benchmark study for a lightweight and low-cost dementia screening tool. The tool is easy to administer, requires no additional experimentation material, and automatically evaluates and indicates potential subjects with dementia. The protocol foresees that subjects answer four distinct tasks, three of which are ordinary questions and one is a counting prompt. In our care use case, older people are assessed remotely via the tool, potentially even via telephone or within a daily care service routine. The assessment results are subsequently sent to professionals who initiate further steps. A machine learning classifier was trained on the French Dem@Care corpus. Solely utilizing vocal features, the classifier reaches 89% accuracy. Implications for the use case and further steps are discussed.

Author Keywords
Dementia, Screening, Vocal features, Classification, Speech analysis, Clinical phonetics

ACM Classification Keywords
J.3.2 Computer Applications: Life and Medical Sciences

INTRODUCTION
Alzheimer's disease (AD), upon other neurodegenerative diseases, has a significant economic impact on our society: according to the World Alzheimer Report 2016, AD is about to become a *trillion dollar disease* by 2018 [21]. This imposes not only a burden on governmental health care budgets, but also has a striking financial effect on private households. As there is no cure to AD yet, research agrees, that early detection of AD is the key to decelerating the disease's progression [2]. The importance of screening — and eventually diagnosis — has been emphasised for the benefits of proactive comprehensive care management, preventing crises and urgent interventions [4].

PervasiveHealth '17, May 23–26, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-6363-1/17/05…$15.00
https://doi.org/10.1145/3154862.3154915

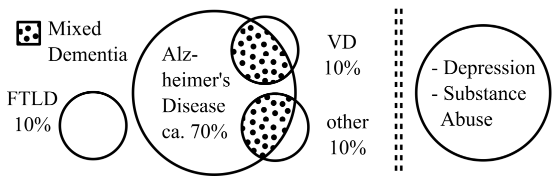

Figure 1: The left panel shows the types of Dementia, according to their cause, including Fronto-Temporal Lobar Degeneration (FTLD), and Vascular Dementia (VD); the dotted areas indicate those cases where more than one cause underlies the disorder. The right panel shows other, mostly reversible, causes for Dementia-like symptoms.

Currently, only around 50% of cases are diagnosed in high income countries, and less than 10% in low and middle income countries, respectively [21]. Hence, major efforts should be spent on identifying criteria for AD preceding obvious mid-stage symptoms, e.g., memory loss, and finding a simple and inexpensive test for broad screening. The main observable neuro-cognitive disorder associated with AD is Dementia: characterised by a decline of cognitive functions, e.g., memory, language, or problem solving, which interfere with occupational or everyday activities, Dementia is typically diagnosed by behavioural patterns, whereas the underlying cause of it is diagnosed by a combination of behavioural and neurological patterns. Causes include AD, as the most important one, vascular disorders, such as strokes, brain tumours, traumatic brain injuries, or fronto-temporal lobe degeneration (FTLD); especially in older persons, there is a high prevalence of AD combined with other disorders, mainly vascular diseases, which is referred to as Mixed Dementia (MD) [25], see also Figure 1.

Diagnosing Dementia and the underlying cause typically includes three steps: first, case history — assessed directly with the patient and also a close third party, second, screening for behavioural markers — the neuropsychological assessment, and third, screening for biological markers. Whereas the screening for biological markers is very costly and strenuous for the patient, the screening for behavioural markers is mainly time consuming. Recently, tools have been developed making the neuropsychological screening for AD more feasible: multiple one-page screening tools which take between

10 and 20 minutes to administer, e.g., the Montreal Cognitive Assessment (MoCA), or the Dementia-Detection (DemTect).

Conducting and evaluating screenings like the MoCA requires the respective material and the physical presence of a professional for manual evaluation. Such screening typically is only conducted after an older person has become clinically apparent; this can be hospital admittance due to neuropsychologically relevant behaviour, neurological incidents demanding emergency care, as well as in the context of a geriatric check-up, or even general practitioners' visit. By nature this does not affect those people who are not being screened due to their low clinical profile, or residence in remote areas with no access to experts.

Therefore, the challenge is to develop a lightweight and low-cost dementia screening method, which (a) is administrable in a few minutes, (b) requires no experimentation set-up nor respective material, and (c) automatically evaluates and indicates potential clinically relevant persons. This would be beneficial due to multiple aspects: (a) little time-consumption, (b) no paper pencil documentation or subject's interaction with additional material and therefore no professional supervision needed, and (c) remotely applicable and fully scalable.

This paper is structured as follows: firstly, we present the opportunities of pervasive solutions in this field and the related work, secondly we deduct our primary use case, then we report the experiment providing evidence for the use case's feasibility and eventually we discuss the results and their implications.

RELATED WORK

Recently, significant progress has been made in the field of computer-supported speech-based AD detection; some researchers even report a completely automated speech-based screening pipeline yielding significant discrimination results [27]. To better understand such research results' applicability for the proposed setting, a careful analysis of the reported experiments is needed; important aspects are (1) the discrimination's targeted groups — especially in terms of assessment interface, (2) the used corpus and tasks, and (3) the type of features extracted from spoken language, i.e., linguistic vs. vocal features. All hereunder reported experiments have been based on natural language data, computer-supported feature extraction from the audio signal or transcripts, and machine-learning classifiers.

The related work can be separated in authors reporting classifiers discriminating between (1.a) cognitive impairment (CI) in a broader sense and healthy controls (HC) with 90% accuracy [30], (1.b) AD and HC with 80% to 95% accuracy [13, 7, 17, 14, 1], (1.c) MCI and HC with 79% to 88% accuracy [14, 27, 10] and (1.d) MCI and AD with 80% [14]. Although a simple broad screening might be binary in nature, e.g., discriminating between AD and HC, for best supporting subsequent interventions, a more comprehensive analysis is needed, compare also Section The Element Care Use Case.

Different corpora have been used in this field to conduct computational experiments. Corpora, such as the *Dementia-Bank* [16], or the *Dem@Care* project's corpus [12], comprise speech samples, partially also transcripts, of mentioned target groups performing different neuropsychological assessment tasks. Such Corpora comprise multiple tasks which means that speech samples used for computational experiments are highly constrained in this sense. Typically, tasks originate from existing validated test batteries, such as the *Wechsler Memory Scale* (WMS), or the *Boston Diagnostic Aphasia Examination* (BDAE), e.g., the famous *Cookie Theft Picture Description Task* which has been the basis for multiple experiments [8, 7, 1, 18], is part of the BDAE. Overall, one can categorise reported work into using (2.a) conversations, e.g., clinical interviews [11], or normal conversations [6, 13], (2.b) free speech tasks, e.g., the CTPDT [7, 1, 14, 18] or descriptions of presented video material [10], (2.c) reading or repetition tasks, e.g., simple sentence reading [17], (2.d) fluency tasks, i.e., phonetic fluency, and semantic fluency [14, 30], as well as (2.e) working memory tasks, e.g., counting backwards [14].

Regarding the above mentioned focus on (a) little time-consumption, (b) no interaction with additional material, and (c) remote applicability, screenings based on free conversations represent the most natural setup. However, for such data, considerable effort has to be spent on preprocessing data, e.g. annotating turns, or trimming the audio file, in order to prepare it for further computational learning and classification. Tasks, eliciting spontaneous speech, are slightly more restricted and therefore easier to process. However, for remote applications visual speech-eliciting material, such as the Cookie Theft Picture, should be avoided for obvious reasons. Using standardised questions/ verbal prompts for eliciting might represent a powerful solution [14, 24], as they could potentially be used via phone. Although less associated with classic speech pathology, but more sensitive for working memory decline, executive tasks, e.g., counting backwards, have high predictive power when it comes to AD vs. HC classification scenarios. This is in line with the ongoing discussion, whether speech-related symptoms — mainly vocal features — are rooted in other language independent cognitive functions, e.g., pauses in counting backwards are related to executive functions or working memory [9].

Some tasks allow for semantic and pragmatic feature extraction, such as conversations and spontaneous speech tasks, as mentioned above, whereas all mentioned tasks allow for phonetic feature extraction.

(3.a) Phonetic features, e.g., hesitations, or pauses, can represent retrieval difficulties [26] indicating the AD related amnestic syndrome of hippocampal type [23]; neuroimaging proves the relation of pauses during autobiographic interviews and episodic memory processes [20]. As AD typically manifests very early in hippocampal brain regions, phonetic features are powerful for discriminating between MCI, early-phase AD, and HC [11, 22, 24]. Spontaneous speech also allows to extract (3.b) semantic and pragmatic features. There is evidence that syntactic impairment in the CTPDT, can be a valuable discriminator [7]. Using the same task, metrics on coordinated, subordinated, and reduced sentences, as well as

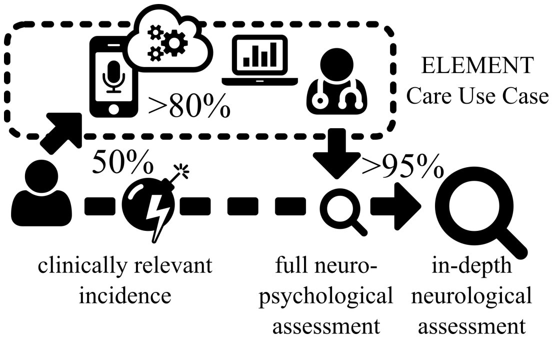

Figure 2: Lower part: traditional sequence of dementia detection; only 50% are detected due to clinical attention, whereas the subsequent assessment has a close to 100% sensitivity and funnels them into full neurological assessment and adequate intervention. Dashed frame: the care use case which targets the detection phase, potentially increasing the number of clinically apparent persons; screening results are reported to a professional, who proceeds in the traditional way.

the use of predicates, and sentence dependencies are reported as powerful features [18].

THE ELEMENT CARE USE CASE

According to the initially stated need for inexpensive broad screening methods, the ELEMENT project aims at broad screening of older persons in professional or informal care settings. Thereby we aim at detecting persons with a highly probable clinical profile and increase the number of persons fed into more comprehensive neuropsychological assessment and intervention by professionals — especially for the benefits of proactive comprehensive care management, potentially preventing crises and urgent interventions [4].

Systematic reviews of computer-supported cognitive screening tools for older people, reveal several key aspects to keep in mind. Given a low IT literacy in an older population, computerised tests may be significantly confounded by interaction artefacts; significant training effects after repetition — older persons' performance might increase due to the fact that they got more used to the novel IT interface — might be mis-interpreted for a treatment effect, or low baseline test scores might be mis-attributed to low cognitive abilities [29]. Therefore, the interface should be un-intrusive and everyday feasible, by avoiding artificial experimental tasks where possible; complex interactions with the IT interface during the tests should be also avoided. Similar to memory functions, symptomatic language deficits are noticeable early on in AD's progression [26]. In this context, measures, such as articulation, word-finding ability, or verbal fluency, have been proven to be powerful indicators for detecting related dementia [28]. In contrast to classic episodic memory tasks, language function tasks, do not comprise a delayed recall, which allows to administer the assessment in a shorter amount of time and therefore more efficiently. Although computerised tests are often associated with the benefit of standardised administration or even complete self-administration, common available test vary a lot in the amount of needed supervision/ guidance,

	AD	HC
n	68	47
Age	78.9 (± 6.5)	72.4 (± 7.9)
Gender	38 F / 30 M	40 F / 7 M

Table 1: Demographics of subjects.

especially when it comes to older persons with cognitive disorders [29]. In this respect, one has to evaluate on a case-to-case basis whether promising fully automated classification pipelines are actually applicable in real life care situations.

Performing quite impressive on binary classification, computerised tests often lack qualitative data a full neuropsychological examination offers [29]. Therefore, the above discussed automatically extracted speech features represent valuable information for professional, going beyond risk assessment. Our care use case foresees the following: Older persons would be asked to answer simple questions, such as telling a short episode in someones life, or performing simple backward counting. This would be partially facilitated and automatically recorded on mobile devices by formal or informal caregivers in care institutions, such as nursing homes, or in the context of informal and formal ambulant home care. The audio signal would be encrypted, anonymised on the device, then analysed on a server and only the result would be sent to a professional, who would either perform traditional further neuropsychological assessment or transfer the person to a respective facility. In case of clinical irrelevance the suggested screening can be repeated on half-year basis; this is far more feasible than traditional screenings, as this use case relies on personal episodic stories only and not on psychometric tests which require costly parallel versions for retesting. Thereby, our envisioned screening application would drastically increase the number of detected and as a consequence comprehensively assessed and treated persons having dementia — compare also Figure 2.

The following section will present an experiment in line with the mentioned use case and the methodological state-of-the-art, serving as first benchmark of our intended use case.

EXPERIMENTAL SETUP

Data

The corpus used consists of 68 samples from older persons, either diagnosed with AD or a form of mixed dementia (MD) which includes AD, and 47 HC; for demographics also see Table 1. All speech samples are in French. The samples include repetition-, denomination-, picture-description-, verbal- and semantic-fluency-, counting- and story-telling-tasks; the data have been collected in the context of the Dem@Care [12] project. According to our use case, i.e., remote applicability with no additional material needed, only a subset of these tasks is used for classification:

Counting Backwards

Participants are asked to count backwards from 305 to 285.

Positive Story

Participants are asked to retell a positive biographic event.

Negative Story
Participants are asked to retell a negative biographic event.

Episodic Story
Participants are asked to retell what they did the day before.

Feature Extraction
For use in classification, we restrict ourselves to vocal features only. Although other results have shown, that using textual information can be beneficial in identifying AD [7] the extraction of transcripts through automated speech recognition (ASR), especially for elderly people, is error prone and would be limited with respect to a scalable language independent system. However, other studies also achieved competitive results using solely vocal features [14, 24]. For each task we extract the following vocal features:

Silence/ Sounding Segments
Determined based on intensity, calculated from the bandpass filtered sound signal, as provided by the Praat [3] software. We manually adjusted the threshold setting between 25 and 28 dB, the minimum silence segment length between 0.25 and 0.5 seconds, depending on the task. The minimum sounding segment length was chosen as 0.1 seconds. We use the maximum, mean, standard deviation, ratio of means of silence and sounding segments' length; the count of silence and sounding segments; the ratio of means of silence and sounding segments' durations and the total silence segments', sounding segments' and tasks duration; the ratio of the length of the maximum sounding segment and the articulation rate; the mean pause duration in the first, second, and last third of the recording.

Voiced/ Unvoiced segments
This is determined by the pitch contour provided by the Praat [3] software. We use the maximum, mean, standard deviation, ratio of means of voiced and unvoiced segments' lengths; the count of voiced and unvoiced intervals; the ratio of mean voiced and unvoiced durations; and the total duration of voiced and unvoiced segments.

Phrase Positions
This is estimated based on sounding segments delineated by silences. We use the position and duration of the first nine phrases in seconds.

Syllable Information
Syllable nuclei are computed in accordance with [5]. We compute the count of syllables, an approximation of the speech rate, as the number of detected syllables divided by the total task duration, and an estimation of the articulation rate, as the number of detected syllables divided by the total duration of sounding segments.

Task Selection
As argued before, we want to provide a feasible assessment tool for a care service scenario. To build such a tool, we need to determine a subset of our recorded tasks that can both perform well and be conducted in concise time. Therefore, we restrict ourselves to tasks that do not rely on any additional material and be can easily performed remotely. We chose the three story telling tasks (positive, negative and episodic) and the counting backwards task, since they are easy and natural to perform and known to have a high predictive power, when discriminating between AD and Control group [14].

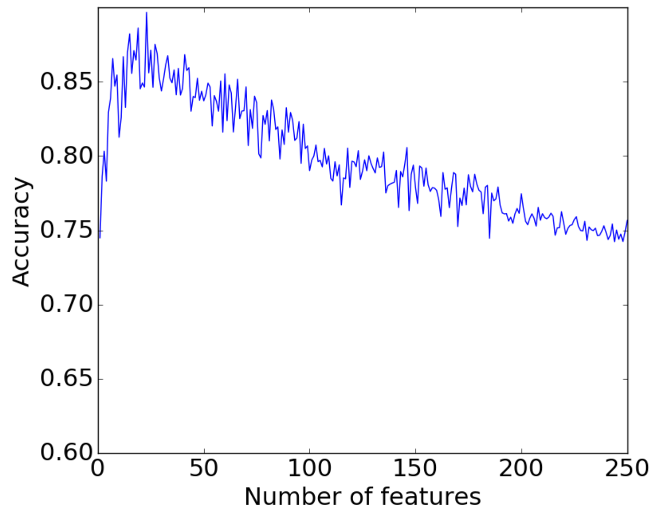

Figure 3: Accuracy (x axis) vs. number of features (y axis) selected by univariate feature selection.

CLASSIFICATION
Due to our limited amount of data — 115 samples — we can not keep a separate hold-out set for testing and instead use k-fold cross validation. In this procedure the data is split into k equally sized folds. The classifier is trained on $k-1$ of them and evaluated on the kth one. We perform a standard 10-fold cross validation. To find a well-performing subset of features, we use univariate feature selection based on mutual information, as part of the classification pipeline.

Performing cross validation on small data sets only once, leads to performance fluctuations between different iterations. To work around this problem, we performed the cross validation 1000 times consecutively and then calculated the mean of all performance metrics.

We use a support vector machine (SVM) classifier with a radial basis function kernel, as implemented by the *scikit-learn* framework [19]. The error parameter C has been empirically determined to be 2 according to the performance. The highest accuracy of 89% ($\pm$ 3) is achieved for a set of 23 features, as seen in Figure 3.

DISCUSSION
This paper set out to provide first results on a low-cost dementia screening concept, which (a) is administrable in a short amount of time, (b) requires neither experimentation set-up nor respective material, and (c) automatically evaluates and indicates potential clinically relevant persons. The chosen tasks, i.e., positive, negative, and episodic story, as well as counting backwards, (a) take approximately three minutes (mean over all 115 test samples = 140s), (b) are entirely speech based and require just a simple question/ instruction,

and (c) are remotely administrable, as they could be fully embedded into an ordinary phone call. The classification results reported are very promising, i.e., an accuracy of 89%, which is of course limited by the small sample size; the corpus comprises over ten tasks and only the four less standardised/ formal of them were chosen according to the use case. Given the reported literature, a classifier with an accuracy of 89% is within the range of state-of-the-art for discriminating between AD and HC. Most importantly, the three story telling tasks chosen, i.e., positive, negative and episodic story, can be repeated in relatively short intervals, without a big learning effect to be expected; one can expect that, for example, a couple of days would be enough for a person to have experienced new positive and negative stories. As early detection of AD-related dementia typically relies on deterioration of cognitive functions over time, this can only be detected by directly comparing test results within one patient. Therefore, the facile repeatability of our task set, is very beneficial, also due to the fact, that we do not need costly parallel test versions for retesting.

FUTURE WORK

Although promising, these are preliminary results. Future work should aim at classification accuracy comparable to existing well-established neuropsychological tests i.e., around 99% sensitivity and over 90% specificity, given an equivalent feasibility. This can be achieved by further optimising the vocal feature extraction, i.e., improving the reliability for detecting silence/ sounding, as well as voiced/ unvoiced segments. By including an ASR, semantic and pragmatic features could be also extracted. There is evidence, that this not only enables a better accuracy [7, 15], but also a more comprehensive screening [26, 22]. However, the remote applicability and scalability should be always taken into consideration, answering the demand for a broad light-weight screening tool for dementia.

ACKNOWLEDGMENTS

This research was partially funded by the EIT Digital Wellbeing Activity 17074, *ELEMENT*. The data was collected during the FP7 *Dem@Care* project.

REFERENCES

1. Al-hameed, S., Benaissa, M., and Christensen, H. Simple and robust audio - based detection of biomarkers for Alzheimer' s disease. In *7th Workshop on Speech and Language Processing for Assistive Technologies* (2016), 32–36.
2. Alzheimer Association. 2016 Alzheimer's Disease Facts and Figures. Tech. Rep. 4, 2016.
3. Boersma, P., and Weenink, D. Praat, a system for doing phonetics by computer. `http://www.fon.hum.uva.nl/praat/`. Accessed: 2017-03-19.
4. Borson, S., Frank, L., Bayley, P. J., Boustani, M., Dean, M., Lin, P.-J., McCarten, J. R., Morris, J. C., Salmon, D. P., Schmitt, F. A., et al. Improving Dementia Care: The Role of Screening and Detection of Cognitive Impairment. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association 9*, 2 (2013), 151–159.
5. de Jong, N. H., and Wempe, T. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods 41*, 2 (2009), 385–390.
6. Dodge, H. H., Mattek, N., Gregor, M., Bowman, M., Seelye, A., Ybarra, O., Asgari, M., and Kaye, J. A. Social Markers of Mild Cognitive Impairment: Proportion of Word Counts in Free Conversational Speech. *Current Alzheimer research 12*, 6 (2015), 513–519.
7. Fraser, K. C., Meltzer, J. A., and Rudzicz, F. Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease 49* (2016), 407–422.
8. Fraser, K. C., Rudzicz, F., and Hirst, G. Detecting late-life depression in Alzheimer's disease through analysis of speech and language. In *Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (2016), 1–11.
9. Goral, M., and Conner, P. S. Language Disorders in Multilingual and Multicultural Populations. *Annual Review of Applied Linguistics 33* (2013), 128–161.
10. Gosztolya, G., Tóth, L., Grósz, T., Vincze, V., Hoffmann, I., Szatloczki, G., Pókáski, M., and Kálmán, J. Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (2016), 107–111.
11. Hoffmann, I., Nemeth, D., Dye, C. D., Pákáski, M., Irinyi, T., and Kálmán, J. Temporal parameters of spontaneous speech in Alzheimer's disease. *International Journal of Speech-Language Pathology 12*, 1 (2010), 29–34.
12. Karakostas, A., Briassouli, A., Avgerinakis, K., Kompatsiaris, I., and Tsolaki, M. The Dem@Care Experiments and Datasets: a Technical Report. Tech. rep., Centre for Research and Technology Hellas (CERTH), 2014.
13. Khodabakhsh, A., Yesil, F., Guner, E., and Demiroglu, C. Evaluation of Linguistic and Prosodic Features for Detection of Alzheimer's Disease in Turkish Conversational Speech. *EURASIP Journal on Audio, Speech, and Music Processing 9* (2015), 1–15.
14. König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P. H., and David, R. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 1* (2015), 112–124.

15. Lehr, M., Prud'hommeaux, E., Shafran, I., and Roark, B. Fully automated neuropsychological assessment for detecting Mild Cognitive Impairment. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (2012), 1039–1042.

16. MacWhinney, B., Fromm, D., Forbes, M., and Holland, A. AphasiaBank: Methods for studying discourse. *Aphasiology 25*, 11 (2011), 1286–1307.

17. Meilán, J. J. G., Martínez-Sánchez, F., Carro, J., López, D. E., Millian-Morell, L., and Arana, J. M. Speech in alzheimer's disease: Can temporal and acoustic parameters discriminate dementia? *Dementia and Geriatric Cognitive Disorders 37* (2014), 327–334.

18. Orimaye, S. O., Wong, J. S.-M., and Golden, K. J. Learning Predictive Linguistic Features for Alzheimer's Disease and related Dementias using Verbal Utterances. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (2014), 78–87.

19. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.

20. Pistono, A., Jucla, M., Barbeau, E. J., Saint-Aubert, L., Lemesle, B., Calvet, B., Köpke, B., Puel, M., and Pariente, J. Pauses during Autobiographical Discourse Reflect Episodic Memory Processes in Early Alzheimer's Disease. *Journal of Alzheimer's Disease 50*, 3 (2016), 687–698.

21. Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., and Karagiannidou, M. World Alzheimer Report 2016 Improving Healthcare for People living with Dementia. Coverage, Quality and Costs now and in the Future. Tech. rep., 2016.

22. Roark, B., Mitchell, M., Hosom, J. P., Hollingshead, K., and Kaye, J. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech and Language Processing 19*, 7 (2011), 2081–2090.

23. Sarazin, M., Chauviré, V., Gerardin, E., Colliot, O., Kinkingnéhun, S., De Souza, L. C., Hugonot-Diener, L., Garnero, L., Lehéricy, S., Chupin, M., et al. The amnestic syndrome of hippocampal type in Alzheimer's disease: an MRI study. *Journal of Alzheimer's disease 22*, 1 (2010), 285–294.

24. Satt, A., Hoory, R., König, A., Aalten, P., and Robert, P. H. Speech-based automatic and robust detection of very early dementia. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (2014), 2538–2542.

25. Schneider, J. A., Arvanitakis, Z., Bang, W., and Bennett, D. A. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology 69*, 24 (2007), 2197–2204.

26. Szatloczki, G., Hoffmann, I., Vincze, V., Kalman, J., and Pakaski, M. Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Frontiers in Aging Neuroscience 7*, 195 (2015), 1–7.

27. Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., Zsura, F., Pákáski, M., and Kálmán, J. Automatic detection of mild cognitive impairment from spontaneous speech using ASR. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (2015), 1–5.

28. Weiner, M. F., Neubecker, K. E., Bret, M. E., and Hynan, L. S. Language in Alzheimer's disease. *Journal of Clinical Psychiatry 69*, 8 (2008), 1223–1227.

29. Wild, K., Howieson, D., Webbe, F., Seelye, A., and Kaye, J. Status of computerized cognitive testing in aging: A systematic review. *Alzheimer's & Dementia 4*, 6 (2008), 428 – 437.

30. Yu, B., Quatieri, T. F., Williamson, J. R., and Mundt, J. C. Cognitive impairment prediction in the elderly based on vocal biomarkers. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (2015), 3734–3738.