Word Repetition in Separate Conversations for Detecting Dementia: A Preliminary Evaluation on Data of Regular Monitoring Service

Kaoru Shinkawa, MS, Yasunori Yamada, PhD
IBM Research - Tokyo, Tokyo, Japan

Abstract

For detecting early signs of dementia, monitoring technology has been actively investigated due to the low diagnostic coverage as well as the requirement for early intervention. Although language features have been used for detecting the language dysfunctions resulting from dementia in neuropsychological tests, features that can be extracted by regular conversations remain unexplored. Here, we propose a feature to characterize the atypical repetition of words on different days which is observed in patients with dementia. We tested it on data obtained from a daily monitoring service for eight elderly people, including two who had been diagnosed with dementia. We found that our feature outperformed the existing linguistic features used in previous studies, such as vocabulary richness and repetitiveness, in terms of effect size and AUC score. The results suggest that the use of our proposed feature holds promise for improving detection performance in everyday situations such as regular monitoring.

Introduction

As the worldwide elderly population increases, the incidence of dementia and Alzheimer’s disease (AD) is becoming an increasingly serious health and social problem. Alzheimer’s Disease International estimated 7% of the world’s population over 65 years old has AD or a related dementia. It also reported that the worldwide cost of dementia may be as high as 605 billion USD a year, equivalent to 1% of the entire world’s gross domestic product. Among the participating countries, Japan is one of those facing a severe aging problem and thus its problems in this regard are very serious. The prevalence of dementia for persons 65 years or older is estimated at around 15%. The cost needed to address this problem in Japan is very high and reaches around 120 billion USD (14.5 trillion JPY). This corresponds to 3% of the entire Japanese gross domestic product. These figures have led to an increasing focus on dementia in recent years. In particular, early diagnosis and intervention has been increasingly recognized as a possible way of improving dementia care, because of recent failures in both clinical trials and laboratory work in the stages of AD. However, diagnostic coverage worldwide remains low. Even in high-income countries, only 40-50% of people afflicted with dementia have received a diagnosis. For example, Kotagal et al. suggested that 55% of people afflicted with dementia do not receive clinical cognitive evaluations in the United States. Monitoring technology capable of detecting early signs of dementia and AD in everyday situations has great potential for supporting earlier diagnosis and intervention. Although there are already several projects and services for monitoring the health of elderly persons with frequent data collection by using mobile applications, whether and how we can exploit the data collected on a daily basis to detect dementia has been largely unexplored.

To detect the health state of elderly persons afflicted with dementia in everyday situations, one of the most prominent candidates is identifying the evolution of language change over the course of dementia and AD. While memory impairment due to shrinking of the medial temporal lobe is the most typical symptom of dementia, both retrospective analyses and prospective cohort studies have shown that language problems are prevalent dating from presymptomatic periods. In addition, language dysfunctions at the time of diagnosis from pathologically proven AD patients with postmortem examination have also been reported. Ahmed et al. showed that they exhibited significant language changes such as syntactic simplification and impairments in lexico-semantic processing. On the basis of these findings, in previous computational work much attention has been paid to explore useful features that can be extracted from data gathered while participants performed neuropsychological tests by professionals such as medical doctors. They have found that the different stages of dementia and AD exhibit specific patterns of language changes in different domains such as phonetics and phonology, morphology, lexicon and semantics, syntax, and pragmatics. Recently, studies on investigating whether these language dysfunctions observed in mental illness including dementia could be extracted under conditions close to those of everyday life have started and have garnered increas-
Although daily monitoring makes it possible to extract language features at one time, comparing language patterns on different days is another possible approach but remains largely uninvestigated.

In this study, we proposed a language feature of dementia that can be extracted by comparing language patterns on different days, and investigated whether it is useful for detecting dementia. To this end, we used conversational data obtained from a regular monitoring service for elderly people in Japan. We investigated the feature using real conversational data provided by a regular monitoring service, and showed how our proposed feature could be useful for detecting dementia.

**Related Work**

In this section, we will show related work that assisted us in aiming to determine the sort of features that can be extracted from conversations through regular monitoring. To this end, we will describe the acoustic and linguistic features in previous studies on speech data where individuals performed neuropsychological tests.

Acoustic features have been widely used for quantifying the individual’s state including emotion, stress, and neurodegenerative diseases such as depression and dementia. One reason for this is that these acoustic features are relatively easy to obtain compared with linguistic features requiring speech-to-text. In the context of detecting dementia, acoustic features have been used for analyzing verbal fluency tasks or speeded-up word-list generation. Previous studies reported that patients with dementia tend to increase their periods of silence as well as the number of pauses they use.

The short-term memory loss associated with dementia makes ordinary conversation difficult because of language dysfunctions such as word-finding and word-retrieval difficulties. These language dysfunctions have typically been characterized by using linguistic features. Picture descriptions were often used for these features and there are several large datasets for speech data in picture description tasks, such as DementiaBank. Here we will provide a brief description of linguistic features widely used in these tasks, because the linguistic features of spontaneous speech used in them could be useful for characterizing everyday conversations.

Syntactic complexity is closely associated with the incidence of dementia. For example, Kemper et al. showed that AD accelerates age-related deterioration in syntactic complexity compared with healthy controls. Syntactic complexity was measured in various ways, such as the mean length of sentences, “part-of-speech” frequency, and dependency distance. Dependency distance infers the number of intervening words between two syntactically related words in a sentence. Another category that must be considered is vocabulary richness. This category measures lexical diversity, which tends to reduce in dementia cases. It was calculated by three typical measures: type-token ratio (TTR), Brunet’s index (BI), and Honoré’s statistic (HS). TTR compares the total distinct word types to the total word count as TTR=U/N. Using the same U and N, BI is also defined as BI=U−Nuni/N. Unlike other measures related to vocabulary richness, for this measure the vocabulary richness becomes greater as BI becomes smaller. HS gives particular importance to unique vocabulary items used only once, also known as hapax legomena. HS is defined as HS=100 log N/(1−Nuni/U). A third feature category is repetitiveness. Some matrices measure the frequency of repeated words and phrases, and others estimate sentence similarities by calculating the cosine distance between two sentences. Another feature category, one widely used in image description tasks, is semantic density. It was calculated on the basis of “informational units” that are predefined objects or text segments that might refer to important information. For example, in the Boston Cookie Theft picture description task, information units consist of objects such as “Woman”, “Cookies”, and ”Boy taking the cookie”. With the information units, semantic density can be defined as the number of information units divided by the total number of words. A number of previous studies have found that individuals with dementia tend to produce speech with lower information, defined as semantic density, than with healthy controls. Semantic density would be difficult to use in daily conversations, because it requires predefined information units.

With regard to language dysfunctions observed in conversations in everyday situations, a previous study based on family reports pointed out atypical word repetition. While this repetition was typically reported to occur in the same
conversation, it also appears in separate conversations that may be held on different days\(^\text{45}\). Repetition in conversation across a number of days be assumed to represent memory impairment that prevents speakers from remembering recent conversations and degrading their social engagement ability by making them less able to expand on conversation topics. Therefore, in this study, we hypothesize that atypical repetition in conversation could be useful for detecting dementia in regular conversations. An attempt was made to infer that atypical repetition of words in a single conversation involved not only repetitiveness but also vocabulary richness because the repetitions could result in a small number of distinct words being used in a conversation\(^\text{46}\). Thus, it might be possible to capture atypical repetition of words in separate conversations by using repetitiveness and/or vocabulary richness.

**Feature Related to Word Repetition in Conversations on Different Days**

In this study, we focused on repetition in conversation as one of the prominent behaviors of dementia. Previous studies characterized atypical repetition of words by the frequency of repeated words or similarity between the sentences in a single conversation. Although descriptive analysis of AD clinical interviews has reported that repetition may also appear in separate conversations that may be held on different days\(^\text{45}\), features that can be extracted from separate conversations remain largely uninvestigated. In this study, we propose a feature that aims to measure word repetition, especially in conversations on different days as addition to the single conversations on the same day.

We first got pairs of conversational data \(D_i\) and \(D_j\) separated by \(t\) days \((T - M \leq t \leq T + M)\) days. \(H_S\) and \(H_S\) were extracted from \(D_i\) and \(D_j\) by calculating Honoré’s statistic \((H_S)\). \(H_S\) is defined as \(H_S = 100 \log N/(1 - N_{\text{uni}}/U)\), where \(N\) is the total number of words, \(U\) is the total number of distinct word types, and \(N_{\text{uni}}\) is the number of total distinct word types used only once. Next, we defined \(D_{ij}\) as a combined document of \(D_i\) and \(D_j\) and extracted \(H_{S_{ij}}\) as a feature of repetitiveness in conversation on different days. Finally, our feature \(R\) was calculated as follows:

\[
R = w_i H_{S_{i}}^{-1} + w_j H_{S_{j}}^{-1} + w_{ij} H_{S_{ij}}^{-1},
\]

\[
w_i + w_j + w_{ij} = 1.
\]

Weights \(w_i\), \(w_j\) and \(w_{ij}\) are hyper-parameters selected by parameter optimization. \(M\) was set to two days in this study.

**Conversational data from a regular monitoring service**

To test the proposed feature, we used daily conversation data obtained from a monitoring service for elderly people provided by Cocolomi Co., Ltd (http://cocolomi.net/). The purpose of their service is to help children to build a connection with their parent living alone by sharing the daily life information of elderly people, such as their physical condition. The communicator calls elderly people once or twice a week to have a daily conversation for about ten minutes. Each conversation is transcribed in spoken word format by the communicator and sent to the family by email (Figure 1). The spoken words of the communicator are eliminated.

The conversational data for analyzing were collected from eight Japanese people (five females and three males; age range 66-89 years, i.e., 82.37 ± 5.91 years old). Two of them were reported as suffering from dementia from the family. Table 1 shows the duration of the service, the number of report calls, the average duration of each call, and the average word length of each report. In total, 458,738 words were used for the analysis. All reports were written in Japanese. For preprocessing, we performed word segmentation, part-of-speech tagging, and word lemmatization on the conversation data. The words tagged as numerals and symbols were excluded from the analyzing data. For the preprocessing, we used the Japanese morphological analyzer MeCab\(^\text{47}\).

**Results**

We investigated our proposed feature \(R\) with conversational data obtained during the phone calls with the regular monitoring service. The feature aims to measure word repetition, especially in conversations that are separated by \(t\) days in addition to a single conversations conducted within a single day.

First, we investigated whether our proposed feature could be used as a measure for discriminating dementia from controls. The discriminative power was measured by using both effect size (Cohen’s d) and area under the receiver
Figure 1: Overflow of regular monitoring service. A communicator calls an elderly customer once or twice a week, transcripts the conversations, and e-mails the transcripts to family members. In this study we analyzed the conversation transcripts this service provided.

Table 1: Participant data list.

| Status   | Gender | Age  | Start      | End        | No. of calls | Ave. call time Mean (SD) [min.] | Ave. word length |
|----------|--------|------|------------|------------|--------------|---------------------------------|------------------|
| Control  | F      | 75-77| 2015 Mar   | 2017 Apr   | 75           | 11.21 (8.85)                   | 395.13 (124.18)  |
|          | F      | 80-83| 2014 Jul   | 2017 Apr   | 109          | 16.63 (4.47)                   | 734.34 (195.10)  |
|          | F      | 87-89| 2016 Jan   | 2017 May   | 104          | 11.15 (4.46)                   | 418.86 (235.12)  |
|          | M      | 66-70| 2014 Jul   | 2017 Apr   | 133          | 10.62 (2.32)                   | 482.89 (118.95)  |
|          | M      | 78-81| 2014 Dec   | 2016 Mar   | 72           | 12.06 (2.83)                   | 554.69 (119.03)  |
|          | M      | 82-85| 2014 Nov   | 2017 Apr   | 226          | 17.75 (6.29)                   | 572.12 (235.49)  |
| Dementia | F      | 85-86| 2014 Jul   | 2015 Nov   | 40           | 9.29 (2.15)                    | 462.28 (204.12)  |
|          | F      | 88-88| 2014 Jul   | 2014 Nov   | 13           | 7.77 (1.72)                    | 277.94 (151.47)  |

operating characteristic curve (AUC-ROC). For Cohen’s d, the 0.8 effect size can be assumed to be large, while the 0.5 effect size is medium and the 0.2 effect size is small. ROC is a graphical plot that illustrates the diagnostic ability of a binary classifier system that ranges from 0 to 1. We respectively set hyper-parameters $w_i$, $w_j$, $w_{ij}$ and $T$ as 0.125, 0.125, 0.75, and 8, which were selected by exploratory experiments. As a result, we found that the proposed feature $R$ for people with dementia was significantly higher than that of controls ($p < 1.0 \times 10^{-24};$ Figure 2). We also obtained the effect size of 2.68 (95% confidential interval (CI): 2.11-3.25) and the AUC-ROC of 0.97 (Figure 3).

We next compared $R$ with other features extracted from single conversation that were typically used in previous studies: vocabulary richness, sentence complexity, and repetitiveness. As for vocabulary richness, we investigated Type-token ratio, Brunet’s index, and Honoré’s statistic. For sentence complexity and repetitiveness, we respectively employed mean sentence length and sentence similarity. We computed the sentence similarity using cosine distance of sentences defined as TF-IDF (term frequency-inverse document frequency) vectors. Among the six features, we observed significant differences between control and dementia in three features: $R$, Honoré’s statistic, and sentence similarity ($p < 1.0 \times 10^{-24}$, $p < 5.0 \times 10^{-20}$, $p < 5.0 \times 10^{-9}$, respectively; Figure 2). In contrast, mean sentence length, type-token ratio and Brunet’s index had no significant difference between the groups ($p > 0.05$). The proposed feature $R$ showed the best results in terms of effect size and ROC ($d=2.68$, ROC=0.97), followed by Honoré’s Statistic ($d=-1.36$, ROC=0.86), and sentence similarity ($d=0.69$, ROC=0.72) (Figure 3).

Our proposed feature $R$ aims to characterize word repetition, especially in conversations on different days in addition to a single conversation on a single day. We investigated the usefulness of the feature representing the repetition in
Figure 2: Feature distributions for control and dementia in our proposed feature $R$ and the existing five features used in previous studies. Boxes denote the 25th (Q1) and 75th (Q3) percentiles. The line within the box denotes the 50th percentile, while whiskers denote the upper and lower adjacent values that are the most extreme values within Q3+1.5(Q3-Q1) and Q1-1.5(Q3-Q1), respectively. Filled circles show outliers, and squares represent mean values.

conversation on different days. We compared it with the feature extracted from single conversation. Specifically, the former feature was extracted from paired conversational data (i.e. $HS^{-1}$ extracted from $D_{ij}$), while the latter was extracted from single conversational data (i.e $HS^{-1}$ extracted from $D_{i}$ ($D_{j}$)). The features extracted from paired conversations and single conversation both showed significant differences between control and dementia ($p < 1.0 \times 10^{-24}$ for single conversation; $p < 1.0 \times 10^{-24}$ for paired conversation). We also found that the features extracted from paired conversations had larger discriminative power than those extracted from single conversations in terms of effect size and ROC score ($d=1.58$, ROC = 0.86 for single conversation and $d=2.67$, ROC = 0.96 for paired conversation; Figure 4). The results suggest that the feature part extracted from paired conversational data contributes to detection performance of our proposed feature.

As an additional analysis, we investigated the tendency of the proposed feature with different intervening days of paired conversations. Specifically, we compared the discriminative power of the feature extracted from paired conversation by changing $T$ from three to 14 days. To define $R$ for paired conversation, we used weight parameter $w_{ij}$ as 1 and others as 0, because we wanted to investigate the relationship between duration $T$ of the paired conversation and word repetitions. In all $T$ values calculated in this study, the proposed R feature for people with dementia was significantly higher than that of controls ($p < 0.05$; Table 2). The effect size and the AUC values respectively ranged from 1.5 to 2.67 and from 0.87 to 0.96. After they increased in the beginning, they peaked at around $T=8$ (effect size of 2.67, 95%
Figure 3: Comparison of our proposed feature \( R \) with the existing five features used in previous studies. Error bars are 95% confidence intervals.

| Feature type           | Effect size (95% CI) | ROC | Color |
|------------------------|----------------------|-----|-------|
| R (T=8)                | -1.36(-1.65, -1.06)  | 0.86|       |
| Honore's statistic     | 0.69 (0.41, 0.98)    | 0.72|       |
| Sentence similarity    | 0.25 (-0.03, 0.53)   | 0.64|       |
| Mean sentence length   | 0.05 (-0.23, 0.33)   | 0.51|       |
| Type token ratio       | -0.07 (-0.35, 0.20)  | 0.48|       |
| Brunet's index         |                      |     |       |

Figure 4: Comparison between our proposed feature \( R \), extracted from single and paired conversations. (A) Histogram and boxplot for each feature. (B) ROC-AUC scores.

CI: 2.10-3.24; ROC = 0.96) and had a tendency to decline (Table 2 and Figure 5).

Conclusion

In light of the increasing demand for detecting dementia in everyday situations, we focused on word repetition in separate conversations on different days on the basis of a previous descriptive study and proposed a feature to characterize
To test our proposed feature, we investigated conversational data obtained from a regular monitoring service. The data the service provided was collected from eight elderly people, including two dementia patients, for a period of up to 30 months at the maximum.

First, we found that our proposed feature has strong discriminating power and achieved up to 2.68 for effect size of Cohen’s d and 0.97 for AUC-ROC scores. We also compared our proposed feature with other linguistic features such as vocabulary richness and sentence similarities. As a result, our feature $R$ outperformed other features, suggesting that the use of our feature in addition to already existing feature sets has promise to improve detection performance. We also analyzed feature extracted from single conversational data and paired conversational data. The results indicate that features from paired conversations on different days may be more advantageous in increasing discriminative power than extracting them from a single conversation. In addition, the results obtained in changing the intervening days of two separate conversational data suggest that our feature $R$ could be especially useful when extracted from conversations of the specific intervening days.

Our study has several limitations. One of the limitations of this study is the small sample of participants. Another limitation is its specific type of conversational data. We only analyzed the conversations from a regular monitoring service, where all conversations were over the telephone. We need to investigate our proposed feature with conversational data collected from everyday life situations such as face-to-face family conversations. In addition, our feature treated a binary state of dementia. The main reason is that the sample size of participants was too small to classify dementia in terms of its severity. In future work we will need to collect data from a larger number of participants, which will allow us to test whether or not our feature could be extended to score dementia on a scalar or ordinal scale. However, because to the best of our knowledge this is the first study which aims to characterize word repetitions in conversations on different days, we believe that using our feature will help make it possible to provide service for detecting dementia in daily monitoring.

**Acknowledgements**

We sincerely thank A. Kamiyama, Y. Masuda, J. Hayakawa, and K. Cho at Cocolomi Co., Ltd. for providing all of the data used in this study, and appreciate valuable comments and suggestions.

---

**Figure 5:** Effect size and AUC of our proposed feature $R$ with different $T$. (A) Effect size of Cohen’s $d$. Error bars are 95% confidence intervals. (B) AUC-ROC score.
Table 2: Discriminative power between control and dementia for our proposed feature $R$ with different $T$.

| T [days] | Control       |         | Dementia    |         | Effect size (95% CI) | AUC-ROC |
|---------|---------------|---------|-------------|---------|---------------------|----------|
| 3       | 5.31 (0.64)   | 389     | 6.25 (0.36) | 17      | 1.5 (0.99, 2.01)    | 0.95     |
| 4       | 5.35 (0.61)   | 506     | 6.43 (0.56) | 20      | 1.79 (1.32, 2.27)   | 0.95     |
| 5       | 5.34 (0.48)   | 679     | 6.48 (0.70) | 22      | 2.31 (1.85, 2.77)   | 0.93     |
| 6       | 5.38 (0.49)   | 543     | 6.63 (0.66) | 20      | 2.51 (2.02, 3.00)   | 0.96     |
| 7       | 5.48 (0.50)   | 350     | 6.78 (0.65) | 20      | 2.45 (1.93, 2.96)   | 0.95     |
| 8       | 5.52 (0.50)   | 299     | 6.88 (0.62) | 17      | 2.67 (2.10, 3.24)   | 0.96     |
| 9       | 5.49 (0.51)   | 238     | 6.58 (0.63) | 20      | 2.11 (1.59, 2.63)   | 0.92     |
| 10      | 5.44 (0.51)   | 134     | 6.50 (0.54) | 17      | 2.09 (1.48, 2.69)   | 0.93     |
| 11      | 5.39 (0.47)   | 76      | 6.44 (0.58) | 11      | 2.16 (1.38, 2.93)   | 0.93     |
| 12      | 5.30 (0.44)   | 55      | 6.04 (0.31) | 7       | 1.72 (0.83, 2.62)   | 0.92     |
| 13      | 5.40 (0.46)   | 42      | 6.05 (0.29) | 10      | 1.50 (0.70, 2.31)   | 0.88     |
| 14      | 5.45 (0.45)   | 34      | 6.18 (0.57) | 5       | 1.57 (0.51, 2.63)   | 0.87     |

References

1. M. Prince, M. Prina, and M. Guerchet. World Alzheimer report 2013 journey of caring an analysis of long-term care for dementia. london: Alzheimer’s disease international, 2013.
2. A. Wimo and M. Prince. World Alzheimer report 2010: the global economic impact of dementia, 2010.
3. Ministry of Health, Labour, and Welfare 2014. http://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000065682.pdf, Accessed September 21, 2017. Japanese.
4. R. Shikimoto, M. Sado, and M. Mimura. The social costs of dementia in Japan: Focusing on the informal care cost. Brain and nerve, 68(8):939–944, 2016.
5. R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, et al. Toward defining the preclinical stages of Alzheimer’s disease: Recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. Alzheimers Dement., 7(3):280–292, 2011.
6. M. Prince, A. Comas-Herrera, M. Knapp, M. Guerchet, and M. Karagiannidou. World Alzheimer report 2016: improving healthcare for people living with dementia: coverage, quality and costs now and in the future. 2016.
7. V. Kotagal, K. M. Langa, B. L. Plassman, G. G. Fisher, B. J. Giordani, R. B. Wallace, et al. Factors associated with cognitive evaluations in the United States. Neurology, 84(1):64–71, 2015.
8. D. Griol and Z. Callejas. Mobile conversational agents for context-aware care applications. Cognit Comput., 8(2):336–356, 2016.
9. P. Pires, L. Mendes, J. Mendes, R. Rodrigues, and A. Pereira. Integrated e-healthcare system for elderly support. Cognit Comput., 8(2):368–384, 2016.
10. H. S. Kirshner. Primary progressive aphasia and Alzheimer’s disease: brief history, recent evidence. Curr Neurol Neurosci Rep, 12(6):709–714, 2012.
11. D. G. MacKay, L. E. James, and C. B. Hadley. Amnesic HM’s performance on the language competence test: Parallel deficits in memory and sentence production. J Clin Exp Neuropsychol, 30(3):280–300, 2008.
12. M. Van Velzen and P. Garrard. From hindsight to insight–retrospective analysis of language written by a renowned Alzheimer’s patient. Interdiscip Sci Rev, 33(4):278–286, 2008.
13. A. Oulhaj, G. K. Wilcock, A. D. Smith, and C. A. de Jager. Predicting the time of conversion to MCI in the elderly role of verbal expression and learning. *Neurology*, 73(18):1436–1442, 2009.

14. S. Ahmed, C. A. de Jager, A.-M. F. Haigh, and P. Garrard. Logopenic aphasia in Alzheimer’s disease: clinical variant or clinical feature? *J Neurol Neurosurg Psychiatry*, 83(11):1056–1062, 2012.

15. S. Ahmed, C. A. de Jager, A.-M. Haigh, and P. Garrard. Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer’s disease. *Neuropsychology*, 27(1):79–85, 2013.

16. B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Trans Audio Speech Lang Process*, 19(7):2081–2090, 2011.

17. W. Jarrold, B. Peintner, D. Wilkins, D. Vergyi, C. Richey, M. L. Gorno-Tempini, et al. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the ACL Workshop on Computational Linguistics and Clinical Psychology*, pages 27–36, 2014.

18. G. Szatloczki, I. Hoffmann, V. Vincze, J. Kalman, and M. Pakaski. Speaking in Alzheimer’s disease, is that an early sign? Importance of changes in language abilities in Alzheimer’s disease. *Front Aging Neurosci*, 7, 2015.

19. M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. Predicting depression via social media. *ICWSM*, 13:1–10, 2013.

20. M. Faurskov-Jepsen, J. Busk, M. Frost, M. Vinberg, E. Christensen, O. Winther, J. E. Bardram, and L. Kessing. Voice analysis as an objective state marker in bipolar disorder. *Transl Psychiatry*, 6(7):1–8, 2016.

21. L.-S. A. Low, N. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen. Detection of clinical depression in adolescents speech during family interactions. *IEEE Trans Biomed Eng*, 58(3):574–586, 2011.

22. V. Mitra and E. Shriberg. Effects of feature type, learning algorithm and speaking style for depression detection from speech. In *Proc IEEE Int Conf Acoust Speech Signal Process*, pages 4774–4778, 2015.

23. A. Esposito, A. M. Esposito, L. Likforman-Sulem, M. N. Maldonato, and A. Vinciarelli. On the significance of speech pauses in depressive disorders: results on read and spontaneous narratives. In *Recent Advances in Nonlinear Speech Processing*, pages 73–82. Springer, 2016.

24. I. Hoffmann, D. Nemeth, C. D. Dye, M. Pakáski, T. Irinyi, and J. Kálmán. Temporal parameters of spontaneous speech in Alzheimer’s disease. *Int J Speech Lang Pathol*, 12(1):29–34, 2010.

25. A. König, A. Satt, A. Sorin, R. Hoory, O. Toledo-Ronen, A. Derreumaux, et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer’s disease. *Alzheimers Dement (Amst)*, 1(1):112–124, 2015.

26. K. D. Mueller, R. L. Koscik, A. LaRue, L. R. Clark, B. Hermann, S. C. Johnson, and M. A. Sager. Verbal fluency and early memory decline: results from the Wisconsin registry for Alzheimer’s prevention. *Arch Clin Neuropsychol*, 30(5):448–457, 2015.

27. M. Nicholas, L. K. Obler, M. L. Albert, and N. Helm-Estabrooks. Empty speech in Alzheimer’s disease and fluent aphasia. *J Speech Hear Res*, 28(3):405–410, 1985.

28. M. Asgari, J. Kaye, and H. Dodge. Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimers Dement (N Y)*, 3(2):219–228, 2017.

29. G. Kavé and M. Goral. Word retrieval in connected speech in Alzheimer’s disease: a review with meta-analyses. *Aphasiology*, 32(1):4–26, 2018.

30. K. C. Fraser, J. A. Meltzer, and F. Rudzicz. Linguistic features identify Alzheimer’s disease in narrative speech. *J Alzheimers Dis*, 49(2):407–422, 2016.

31. A. Khodabakhsh, S. Kusxuoglu, and C. Demiroglu. Natural language features for detection of Alzheimer’s disease in conversational speech. In *IEEE EMBS Int Conf Biomed Health Inform*, pages 581–584, 2014.

32. P. Garrard, L. M. Maloney, J. R. Hodges, and K. Patterson. The effects of very early Alzheimer’s disease on the characteristics of writing by a renowned author. *Brain*, 128(2):250–260, 2004.
