Inhibitory Control and the production of disfluencies in speakers with Alzheimer’s Disease

Simon Williams, Claire Lancaster, Clea Tanner

University of Sussex, UK
s.a.williams@sussex.ac.uk

Abstract

This paper presents preliminary findings from a longitudinal study that investigates the ability of speakers diagnosed with Alzheimer’s Disease (AD) to manage distractions and the same speakers’ production of hesitations in semi-spontaneous speech. Previous research has found that two kinds of hesitation: silent pauses and reformulations, are significantly influenced by inhibitory control in healthy second language learners. Silent pauses are also well-known linguistic markers of disease progression in people living with AD and may evidence a speaker’s production problems, whereas reformulations promote the joint discourse. Inhibitory control is therefore associated with disfluencies that are personal and social.

It was hypothesised that first language speakers with a diagnosis of AD would make fewer reformulations and produce more silent pauses relative to healthy individuals. Data that included a Stroop task, a two-minute monologue, and a two-minute category fluency task were collected and analysed from 11 participants with AD and 13 healthy Controls.

In AD participants, only the number of silent pauses correlated with the Stroop task, while reformulations correlated with the MoCA, and silent pause duration with the MoCA and category fluency task, suggesting that the effects of disease-related cognitive decline rather than competing information were responsible for the lack of fluency.

Index Terms: Alzheimer’s Disease, disfluencies, inhibitory control, reformulations, semi-spontaneous speech

1. Introduction

The speech of people diagnosed with Alzheimer’s Disease (AD), the most widespread form of dementia, is recognised for its linguistic attrition and markers of disfluency, most ubiquitous of which is the silent pause (SP) (Table 1). A recent review of natural language processing techniques for age-related pathologies of dementia [1] surveys a plethora of linguistic biomarkers so far adopted by researchers. They include paralinguistic (e.g. eye gaze), text-based, and speech-based features. All three categories are porous. For example, although the text-based category refers to the analysis of written data, its object is the lexical, syntactic, semantic, and pragmatic tokens also generally accessible in spoken language. The speech-based studies include an array of acoustic features, such as prosodic temporal properties of speech rate and pause rate, and disfluencies. Linguistic changes in all three categories are considered to be ‘a reliable, well-established source of digital linguistic biomarkers’ [p. 6] and have been associated with changes in ‘episodic and semantic memory [and] executive functions’. Executive functions comprise the set of self-regulatory processes that enable people to plan, prioritise, manage distractions, and focus on goal-setting and task achievement [2]. Several recent studies have reported on the potential of automatic speech recognition to identify AD based on markers in spontaneous speech [3, 4, 5, 6, 7]. Only Gosztolya et al. [4] base their approach partly on linguistic features they take from Hoffmann et al. [8] and that include SPs.

Claims have been advanced for the economical diagnostic potential of the SP with automatic machine reading, despite the potential complicating factors of background noise, multiple speakers in certain speech events, and speaker health characteristics. AD is most common in people over 65 years [9], by which time people may have a number of other pathologies affecting their speech. Examples of speech events in which pauses may be more or less difficult to assign to a single speaker are interview, discussion, and narrative, all of which are commonly used in current diagnostic procedures. The question arises whether alternative disfluency markers might be legitimised and offer their own affordances.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full term</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Alzheimer’s Disease</td>
</tr>
<tr>
<td>CAT</td>
<td>Category fluency task</td>
</tr>
<tr>
<td>MoCA</td>
<td>Montreal Cognitive Assessment</td>
</tr>
<tr>
<td>REF</td>
<td>Reformulation</td>
</tr>
<tr>
<td>SC</td>
<td>Stroop Colour (sub-component)</td>
</tr>
<tr>
<td>SP</td>
<td>Silent pause</td>
</tr>
<tr>
<td>STR</td>
<td>Stroop task comprising SC and SW</td>
</tr>
<tr>
<td>SW</td>
<td>Stroop Word (sub-component)</td>
</tr>
</tbody>
</table>

Working with semi-spontaneous monologues, Korko and Williams [10] found that reduced inhibitory control in 82 second-language speakers of English ($N_{females} = 47$, $M_{age} = 26.35$, $SD_{age} = 6.49$, 19–46 years), as measured by the Stroop task, was associated with SPs and reformulations (REFs) alone among a collection of disfluencies. Inhibitory control, or response inhibition, is an executive function that allows individuals to override prepotent or habitual responses in favour of actions more consistent with achieving their goals [11]. For example, second-language speakers may experience a number of distractions from their first language, increasing the need for inhibitory control to optimise their second-language performance. The effect of interference is to impose a burden on executive functioning in the production of the second language. Analogously, people living with AD when speaking in their first language may experience a reduced ability to maintain thought and language sequencing caused by poor executive functioning and short-term memory.
There is little doubt that the availability of working memory, executive functioning, and access to the mental lexicon, is responsible for many of the SPs observed in speakers with AD. The relationship may be compared to the increasing response latency in a Stroop task when subjects are asked to name the colour ink of words whose semantic meaning differs, e.g. the word ‘blue’ written in a red font. Increased latencies and task errors are indicative of reduced inhibitory control [12], and hence a similar association would probably exist in individuals diagnosed with AD speaking their first language. It was hypothesised that people living with AD would produce (1) more silent pauses and (2) fewer REFs, and that this would correlate with poorer performance on the Stroop-inhibition task. SPs are defined in this study as the absence of speech in the course of an utterance outside that produced for rhetorical effect or for inhalation or exhalation. For example, she was on an island running her. outdoor pursuits centre [0.469] [inbreath 0.557 not counted as an SP] so we spent a week with her in this great. really good resort [0.296] and er [0.685] that was really nice (SPs in seconds within square brackets, ‘.’ indicates pause < 250 ms, this data)

REFs are defined as the speakers’ abandonment of an utterance once begun only to start anew, whether or not the topic of the new start is the same as that of the trouble source (cf. Different repairs in [13]), e.g.

I think we went to I don’t think we’ve been to Germany much (reformulation underlined, this data)

REFs imply concern for discourse clarification and hence addressee understanding, and speakers require additional working memory to simultaneously hold and process the repairable structure and repair.

2. Method

2.1. Participants

All participants were 65 years or older, fluent in English, did not have communication problems, had not had a stroke, and had not experienced a traumatic brain injury. In addition, inclusion criteria for controls stated that they should not have received a diagnosis of AD. Twenty-six participants with English as their first language took part in the study, 13 with AD (N_{female} = 1, M_{age} = 77.4, SD_{age} = 5.98, 69–88 years) and 13 healthy Controls (N_{female} = 10, M_{age} = 74.1, SD_{age} = 5.90, 68–87 years). People living with dementia in the UK are more likely to be women [14], making our all-male sample atypical. Our Control gender is similarly imbalanced in being predominantly female, so it is not possible to say whether gender impacts our results. Participants with AD were recruited from NIHR Join Dementia Research (www.joindementiaresearch.nihr.ac.uk/). Six Controls were the spouses of participants and lived at the same address. Other Controls were recruited from Join Dementia Research or opportunistically through advertisement.

2.2. Materials and procedure

After going through and signing the consent form, participants who were not recruited from Join Dementia Research, where some diagnostic information was available to researchers, were asked to confirm they had received an AD diagnosis and to give its date. Both sources of recruitment involved an element of participant self-report, often supported by the corroboration of carers. Participants were informed of the availability of a rest break after the Stroop task and their responses to all four tasks were audio-recorded.

To help quantify participants’ level of impairment, the 30-item paper format of the Montreal Cognitive Assessment (MoCA) [15], a widely-used screening instrument for cognitive impairment, was included in the protocol. The original MoCA [16] was validated on 90 controls, 94 patients with mild cognitive impairment (MCI) and 93 patients with Alzheimer’s Disease, and shown to be sensitive to a wider range of memory loss than the MMSE, the nearest equivalent. Of the seven cognitive domains that the MoCA claims to measure, short term memory, executive functions, and language were of particular relevance in the present study.

We also administered a category verbal fluency task (CAT) [17] as an alternative language-based measure of cognitive functioning. The task focuses on word production and participants were asked to name as many animals as they could within 2 minutes.

2.2.1. Stroop task

In order to minimise the burden on participants, we administered a paper-based version of the Stroop task [18] (STR) in which participants were first required to read as many as possible of 112 monosyllabic words naming four colours (red, blue, green, tan), randomly written in discordant red, blue, green, and tan coloured fonts (SW). For example, the word ‘blue’ might be written in red ink and participants would be expected to say ‘blue’. The words were formatted in four columns. Participants were then asked to read a similar page of 112 words, this time naming the colour ink that the four words were printed in (SC). For example, the first word on the new page was ‘blue’ and it was written in red ink so participants would be expected to say ‘red’. Participants were given a time limit of 2 minutes in each case.

2.2.2. Speech production task

Participants were handed a card containing the following written prompt, modelled on those in [19] and [20], and asked to speak about it for 2 minutes (Figure 1).

Describe a journey you remember well.

You should say:

• How you travelled
• Where you went
• What you saw

And explain why the journey was memorable for you.

Figure 1: Participant topic card.

Before they began speaking, participants were given one minute to prepare and had the opportunity to make paper and pencil notes. The task administrator was requested not to speak or back-channel (mm-hm etc.) during the participant’s long turn but to supply non-verbal encouragement in the form of nods, smiles, and hand gestures for the participant to continue speaking or to signal when to stop. If the participant ran out of

19
words, the administrator waited for 5 seconds, then pointed to one of the three bullet points in the prompt and said, “Can you tell me a bit more about this?” If the participant indicated they still had nothing further to say or wanted to stop, the administrator sought confirmation and ended the task.

The speech recordings were transcribed and REFs identified manually in the transcripts by one of the authors. The same author used Audacity Version 2.4.2 audio editor to import the MP3 audio files and listened while following the wave form to locate possible silences. Visual identification alone would have been insufficient as some silences were masked by background noise; in other places, what looked like silence was a speaker’s intake of air or an outbreath, and consequently repeated listening at a macro- and micro-level was necessary to establish SPs. Following common practice, only silences of 250 ms or more were counted, as smaller absences of speech may occur from occlusives or the shift between consonants [21]. The number and duration of SPs were separately analysed.

Data collection took place over a three-week period. The majority of participants and Controls were visited in their own homes but one Control chose to come to the University. Data elicitation took approximately 45 minutes.

### 3. Findings

One participant with AD was colour blind and so could not take part in the Stroop task; another with AD did not want to complete the long speech turn. As data from the Stroop task and the long speaking turn were key variables in the study, results from these two participants were excluded from the analysis, leaving data from 11 participants with AD ($N_{\text{male}}=0$, $M_{\text{age}}=77.36$, $SD_{\text{age}}=6.23$, 69–88 years) and 13 healthy Controls as before. Scores for the MoCA distinguished the two groups (Table 2).

#### Table 2: Descriptive statistics for MoCA scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>Variables</th>
<th>$M$</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants with AD</td>
<td>4-26</td>
<td>14.55</td>
<td>7.461</td>
</tr>
<tr>
<td>Controls</td>
<td>23-30</td>
<td>27.00</td>
<td>2.082</td>
</tr>
</tbody>
</table>

The two parts of the Stroop task: reading aloud the coloured words and naming the colour of the font, were scored in the same way. One mark was awarded for each correct answer, giving a possible total of 112 per part. As the task was time-limited, no tally was kept of latency. Two unforeseen problems arose. Some participants did not recognise the word ‘tan’ and often substituted ‘yellow’. They were credited as though they had said ‘tan’. Four other participants had intermittent difficulty distinguishing either blue and green (three Controls) or, in one case, green and tan (participant with AD).

We first report correlation analysis for the association of STR, MoCA and CAT with SPs and REFs in participants with AD and afterwards in Controls. We go on to report differences variable by variable between the two groups. All correlation analyses reported below are two-tailed. In participants diagnosed with AD, Spearman’s rank-order correlations revealed a significant association between STR and SP$_{\text{number}}$ ($r_s = -.715, p = .013$) (Figure 2) and SC and REFs ($r_s = .653, p = .029$) but no significant association between STR and REFs ($r_s = .336, p = .313$) or STR and its sub-components with SP$_{\text{length}}$ ($r_s = -.400, p = .223$) (Table 3). The MoCA showed significance for SP$_{\text{length}}$ ($r_s = -.756, p = .007$), SP$_{\text{number}}$ ($r_s = -.792, p = .004$) and REFs ($r_s = .643, p = .033$). The CAT showed significance for SP$_{\text{length}}$ ($r_s = -.558, p = .027$) and SP$_{\text{number}}$ ($r_s = .626, p = .040$), indicating an association with cognitive function and possibly word searches at least for some SPs. Spearman gave no correlations between any of the variables in Controls.

#### Table 3: Spearman correlations for STR with REFs and SPs.

<table>
<thead>
<tr>
<th>Participant group</th>
<th>Variables</th>
<th>Spearman</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>STR, SP$_{\text{number}}$</td>
<td>-.715</td>
<td>$p = .013$</td>
</tr>
<tr>
<td></td>
<td>STR, SP$_{\text{length}}$</td>
<td>-.400</td>
<td>$p = .223$</td>
</tr>
<tr>
<td></td>
<td>STR, REFs</td>
<td>.336</td>
<td>$p = .313$</td>
</tr>
<tr>
<td></td>
<td>SC, REFs</td>
<td>.653</td>
<td>$p = .029$</td>
</tr>
<tr>
<td>Controls</td>
<td>STR, SP$_{\text{number}}$</td>
<td>.152</td>
<td>$p = .620$</td>
</tr>
<tr>
<td></td>
<td>STR, SP$_{\text{length}}$</td>
<td>-.036</td>
<td>$p = .907$</td>
</tr>
<tr>
<td></td>
<td>STR, REFs</td>
<td>-.026</td>
<td>$p = .934$</td>
</tr>
</tbody>
</table>

Mann-Whitney U tests were next conducted to determine whether there were differences between the AD and Control groups in their production of SPs and REFs. The results indicate a significant difference between groups for SP$_{\text{length}}$ (AD $M=29.7$ s, $SD=28.7$ s, Controls $M=6.51$ s, $SD=3.87$, $U=33.00$, $p = .006$) (Figure 3) though not for SP$_{\text{number}}$ (AD $N=208$, $M=18.9$, $SD=10.2$, Controls $N=151$, $M=11.6$, $SD=6.35$, $U=41.00$, $p = .077$) or REFs (AD $N=41$, $M=3.73$, $SD=3.64$, Controls $N=29$, $M=2.23$, $SD=2.98$, $U=53.00$, $p = .277$).

#### Figure 3. AD and Control group production of SP$_{\text{length}}$. 

Figure 2. Scatterplot correlations of AD speakers’ STR with SP$_{\text{number}}$. 

Figure 3. AD and Control group production of SP$_{\text{length}}$. 

---

**Table 3:** Spearman correlations for STR with REFs and SPs.
Even though one Control was identified as a REF outlier, the case was not extreme and its removal resulted in no significant change ($U=42.50, p = .240$). Performances of the two groups was significantly different on STR ($U=24.50, p = .006$), CAT ($U=17.50, p = .002$) and MoCA ($U=4.50, p = .001$).

4. Discussion

The association in AD participants between better performance on the MoCA and more REFs is consistent with the argument that individuals with higher MoCA scores have stronger executive abilities. This allows them to hold in place the repairable sequence, calculate alternatives, and produce reformulations for clarity and so demonstrate prosocial behaviour. The range of $SP_{length}$ and REFs is much greater in AD participants than it is in Controls. Control group production of $SP_{length}$ is narrower in scale and that of REFs generally narrower and fewer.

The findings suggest that in healthy speakers, the presentation of $SP_{length}$ and REFs is a stylistic feature relating to individual difference and within the control of the speaker in spite of being simultaneously a reactive phenomenon. For speakers with AD, the lack of significance between Stroop and $SP_{length}$ or REFs may not directly reflect inhibitory control. The sample is small and it is worth noting that the study may be underpowered: the number of REFs may be less a matter of individual difference and more probably one of pathology. Although the association of AD speakers' REFs with the results of the MoCA did not translate to a significant difference from Controls for this disfluency, the Controls were not all at ceiling on the MoCA (Table 2), and the MoCA scores of two participants with AD were as high or higher than the lowest Control.

In addition, while all participants with memory problems in this study were recorded on the Join Dementia Research site as having been diagnosed with AD rather than any of the other nine types of dementia listed, e.g. vascular dementia or dementia with Lewy bodies, it is likely that this group of participants presented with a wider range of memory loss types, some of which might have impacted their speech and language in different ways. Join Dementia Research volunteers could only select one option for diagnostic type and had perhaps opted for the more general term. Notwithstanding that the STR test results showed the groups to be significantly different for inhibitory control, the hypothesis that performance on the Stroop-inhibition task of people living with AD would correlate with fewer REFs and more silent pauses was only partly supported by the findings. Only the number of silent pauses and REFs with SC showed such a relationship. Accordingly, we conclude that inhibitory control is not related to the length of SPs or the number of REFs between AD and Control groups. The fact that significance was achieved in AD participants for the association of $SP_{number}$ in relation to the STR, for $SP_{length}$ and REFs in relation to the MoCA results, and for $SP_{length}$ in relation to the CAT indicates that these two disfluencies are impacted by disease progression, resulting in more SPs (number and length) and fewer REFs, rather more than by the ability to inhibit competing information.

It remains to be seen whether, after 12 months and with repeat data elicitation from the same participants, AD speakers' SPs increase in number and/or length and the number of REFs decreases, as suggested by the above; and whether the results of healthy Controls continue to be non-significant in this respect.

5. Conclusions

Although the original premise for the investigation may have been mistaken, the findings so far are not without value. That SPs tend to increase and other disfluences to decrease with progression of AD is well recognised [22]. The absence of significance between STR, $SP_{length}$ and REFs in AD participants’ data probably reflects the small sample size, the heterogeneity of disease type, and the cross-sectional nature of the study so far. In Pistono’s study [23], AD REFs (‘self-corrections’) correlated with lexical-semantic tasks, cf the CAT, and were thought to be associated with word searches. In this study, AD participants’ $SP_{number}$ was significantly correlated with the STR, and $SP_{length}$ with the MoCA and CAT, suggesting that, as implied by the Control data, the prevalence of the two variables, i.e. REFs and SPs, may indicate individual differences that make accessing the mental lexicon more difficult, rather than being related to inhibitory control. The Mann-Whitney U revealed a significant difference between the duration of SPs between the groups, so it seems that the explanation for their marked presence in the AD group is due to something other than lack of inhibitory control, perhaps to difficulties in ordering a sequence of thought and language and not the need to resolve competing information.

The longitudinal nature of the study will test this explanation. We intend to repeat the data collection from the same participants in autumn 2023 to confirm whether the preliminary findings become more pronounced over time. Further data will provide a means of assessing the association over time of executive function, lexical retrieval, and pro-social behavior, as indicated by speech REFs in individuals at various stages of AD. Monologue elicitation is an example of an easy to scale-up diagnostic tool that has the potential to provide economies in early assessment and intervention. Extension of the project might lie in the development of therapeutic interventions that exploit the findings by including, for example, structured activities in which individuals are provided the space to monitor and revise their spoken output, thus making full use of working memory. Examples of provision might include an app that supports the group construction of narratives in which the negotiation of successive phases gives speakers opportunities to clarify their suggestions by using REFs.

6. Acknowledgements

This research was supported by Alzheimer’s Research UK South Coast Network. We would also like to thank the participants and acknowledge Join Dementia Research as a source of participant recruitment. Join Dementia Research is funded by the Department of Health and Social Care and delivered in partnership by the National Institute for Health Research, Alzheimer Scotland, Alzheimer’s Research UK and the Alzheimer's Society. The study received ethical approval from the University of Sussex on 1 September 2022.
7. References


