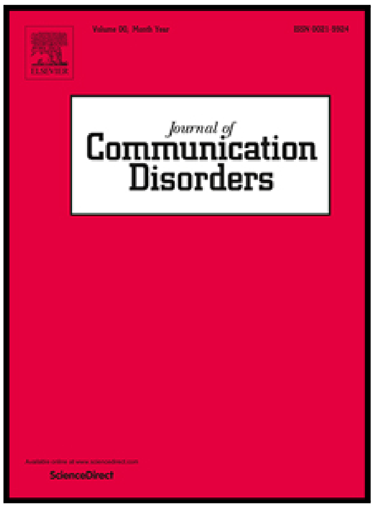
Journal Pre-proof



Breaking the flow of thought: increase of empty pauses in the connected speech of people with mild and moderate Alzheimer's disease

Mary Lofgren , Wolfram Hinzen

PII: S0021-9924(22)00033-8
DOI: https://doi.org/10.1016/j.jcomdis.2022.106214
Reference: JCD 106214

To appear in: *Journal of Communication Disorders*

Received date: 29 January 2021
Revised date: 3 March 2022
Accepted date: 15 March 2022

Please cite this article as: Mary Lofgren , Wolfram Hinzen , Breaking the flow of thought: increase of empty pauses in the connected speech of people with mild and moderate Alzheimer's disease, *Journal of Communication Disorders* (2022), doi: https://doi.org/10.1016/j.jcomdis.2022.106214

Breaking the flow of thought: increase of empty pauses in the connected speech of people with mild and moderate Alzheimer's disease

Mary Lofgren[a], Wolfram Hinzen[a,b]

[a]Dept. Translation & Language Sciences, Universitat Pompeu Fabra,

(Carrer Roc Boronat, 138, Barcelona, Spain 08018)

[b]Intitut Català de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

(Passeig de Lluís Companys, 23, 08010 Barcelona)

This study was conducted with support from the Ministerio de Ciencia y Innovación, grant PID2019-105241GB-I00 and with support from the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the European Social Fund. Additionally, this study made use of the Pitt corpus, funded by the following grants: NIA AG03705 and AG05133.

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca

Unió Europea
Fons social europeu
L'FSE inverteix en el teu futur

Corresponding author:
Mary Lofgren
mary.lofgren@upf.edu

Breaking the flow of thought: increase of empty pauses in the connected speech of people with mild and moderate Alzheimer's disease

Abstract.

Introduction: The profile of spontaneous speech in Alzheimer's disease (AD) includes increased pausing as a window into cognitive decline. We here aimed to further characterize the pausing profile of AD by linking pauses to the syntactic positions in which they appear and disease progression. *Methods*: Speech was obtained through a picture description task, thus minimizing demands on episodic memory (EM), from a group of mild (N=21) and moderate AD (N=19), and healthy elderly controls (N=40). Pauses were sub-indexed according to whether they occurred within-clauses, clause-initially, or utterance-initially, and whether they preceded nouns, verbs, or adjectives/adverbs, when occurring within-clauses. Additionally, relations to verbal fluency (VF) measures at the single-word level were explored. *Results*: Pause rate but not duration distinguished controls from both AD groups, while fillers did not distinguish any groups. The analysis by syntactic position revealed a highly differentiated picture, with largest effect sizes of significant group differences seen in the utterance-initial pause rate. The two AD groups patterned differently when compared to controls, while none of the measures differentiated the AD groups. Specifically, moderate but not mild AD differed from controls in clause-initial pauses, while mild but not moderate AD differed from controls in within-clause positions. At the within-clause level, the effect dividing controls from mild-AD was specifically driven by pauses ahead of nouns. A significant negative correlation emerged between pausing rate in spontaneous speech and VF measures in the mild-AD group only. *Conclusions*: Increased empty (non-filled) pauses in AD are not confined to pauses in within-clause positions, which are most directly related to problems in the retrieval of words. Even in early disease stages, where these within-clause pause effects are seen, they are confined to nouns, revealing a grammatically specific problem possibly related to the referencing of objects. At all disease stages, pauses increase in utterance-sized units of structure, indicating progressive problems in the creative configuration of complete thoughts.

1. Introduction

Decline in episodic memory (EM) is a clinical hallmark of Alzheimer's disease (AD) dementia, but progression of the disease results in decline in other cognitive functions (López-de-Ipiña et al., 2013). One cognitive domain of increasing interest is language (Boschi et al., 2017; Laske et al., 2015). Language deviance is detectable in spontaneous speech long before the criteria for a mild cognitive impairment (MCI) or AD diagnosis are met (Ahmed et al., 2013). A prominent finding on connected speech in AD has been word-finding difficulties, when speakers intend to refer to a particular entity and cannot name it in a precise fashion (Ostrand & Gunstad, 2020). This can reflect decline in semantic memory (SM), which is also evident in early decline of performance on verbal fluency (VF) tasks at the single word level (Henry et al., 2004) and naming tasks (Joubert et al., 2010). A primacy of EM impairment over linguistic impairment would predict alterations in language measures specifically when such memory demands are high. Altered language patterns, however, have been found in non-past-directed spontaneous speech tasks, literary writing, and picture descriptions (Addis et al., 2009; Garrard et al., 2005; Ahmed et al., 2013), which do not require recall of specific and unique episodes from autobiographical memory. SM load, though, definitely remains involved, insofar as lexical concepts have to be retrieved when objects and events are referenced.

A relative independence of language impairment relative to EM impairment in AD is also suggested by the fact that linguistic deviance in AD includes differences in syntactic organization, which are not assumed to reflect impairments in (declarative) memory (Ahmed et al., 2013; Emery, 2000; Liu et al., 2020; de Lira et al., 2011; Orimaye et al., 2017; Chapin et al., 2021). On the other hand, there is also evidence that language and EM interact. Thus, the use of verbal labels influences access to episodic features (Wang & Gennari, 2019). More speculatively, retrieving episodic memories depends on generating a structurally complex mental representation containing the required specifications for the when, where, what, and who of a personally experienced event, which resembles the complexity found in the sentences encoding such memories (e.g., *I had eggs for breakfast with John this morning*). A weakened language system could affect the capacity to assemble such structures, and hence EM performance. Independent evidence in this direction is the fact that working memory (WM) and EM decline is seen in both post-stroke (Lang & Quitz, 2012) and primary progressive aphasia (Ash et al., 2013; Nilakantan et al., 2017; Win et al., 2017). Connections between EM and language are also evident in amnestic individuals with hippocampal damage, who have been found to manifest language deviances in future-directed or scene construction tasks as well, which involve no EM (Duff & Brown-Schmidt, 2012; Warren et al., 2018).

Apart from SM, EM, and WM, it is clear that a full picture of language decline in AD, particularly in spontaneous speech production, will need to include the factor of general processing speed (PS), which is arguably related to WM (Salthouse, 1996) and contributes to VF decline in AD (McDowd et al., 2011). It was also related to naming speed in AD (Warkentin et al., 2008), and to general temporal measures of speech (pausing) in the storytelling of an aphasia sample (DeDe & Salis, 2020). Despite some evidence of the role of both PS and WM in discourse production, however, the relative contributions of these cognitive domains have not been investigated directly (Martin & Slevc, 2014, p. 441). Furthermore, typical studies investigating different types of memory storage and processing have relied on linguistic tasks from which memory measures and processing speed were derived (e.g., naming or discourse production tasks), thus confounding the cognitive measures by the linguistic nature of the task.

Prior to clarifying the exact picture of cognitive factors contributing to language-related decline, more fine-grained insights into the profile of language decline in AD have to be obtained, viewed as something that can inform future inquiries into the cognitive mechanisms involved. We aimed to do this here at the specific level of temporal aspects of fluency in connected speech. Pauses interrupting a continuous speech phonation stream are a universal feature of neurotypical speech, yet alterations in speech pause frequency or duration can reveal anomalous 'gaps' opening in the speech generation process. These have shown sensitivity to forms of cognitive decline across multiple pathologies (Boschi et al., 2017; Çokal et al., 2019b; Gayraud et al., 2011; López-de-Ipiña et al., 2013; Martínez-Sánchez et al, 2013; Mack et al., 2015; Pistono et al., 2016; Roark et al., 2011; Szatloczki et al., 2015).

Importantly, acoustic measures are relatively easy to automatize, with profiles obtainable within seconds from spontaneous speech data using freely available tools (Mertens, 2004). This suggests they could play a role in clinical applications such as disease monitoring over time or early detection.

In AD, only a small number of studies have investigated fluency in connected speech in the past decade using a variety of measures (López-de-Ipiña et al., 2013) and clinical groups. Some studies have target relatively broad measures of speech production, e.g., speech rate, which has been shown to be lower in individuals with AD and MCI (Boschi et al., 2017; Gayraud et al., 2011; Pistono et al., 2016; Szatloczki et al., 2015). Regarding pauses, the overall frequency of pauses produced has been found to be higher in the speech of AD participants than that of healthy controls (Gayraud et al., 2011; Martínez-Sánchez et al., 2013; Pistono et al., 2019; Roark et al., 2011).

Importantly, pause frequency has also been found to be sensitive to task effects. Pistono et al. (2019) directly investigated task-effects by comparing fluency patterns in mild AD participants when performing a memory-based narrative and a picture-based narrative, which imposes minimal EM demands. While pause duration was significantly different between groups on both the picture-based narrative and memory-based narrative, pause *frequency* was only a distinguishing variable in the picture-based narrative. This apparent task effect provides insight into the potential function of pause behavior. Pistono et al. (2019) specifically suggest that increased pause frequency represents a compensatory mechanism, possibly related to lexical retrieval/memory capacities. A positive correlation between semantic fluency test scores at the single word level and the pause rate in the picture description task is provided as evidence for this claim. Interestingly, in the memory-based narrative, pause rate did not correlate with semantic fluency, but did correlate positively to scores on a recall task intended to measure anterograde EM. In short, while higher pause rates in an EM-related task correlated with better scores on EM tests, in a picture-based narrative the pause rates no longer correlated with EM test performance. These findings indicate that there may be no generalized increase in pause production, but that pausing behaviour relates to the specific memory demands of the task. Based on these results, the authors propose that pauses function as a potential compensatory mechanism in the earliest stages of AD, which they suggest should be taken as a positive sign, i.e. speakers are attending to task-specific demands (Pistono et al., 2019).

A critical gap in the current literature on disfluency in dementia lies in the analysis of pauses relative to their syntactic context, which can illuminate the question of their cognitive function. While fluency metrics target language at the relatively 'surface' level of its acoustic organization, they show rich interactions with 'deeper' aspects of language, such as its syntactic organization. On the one hand, there is evidence that pausing patterns reflect syntactic complexity (Ferreira, 1991; McDaniel et al., 2009; Mirdamadi & De Jong, 2015). On the other hand, units of different syntactic complexity correspond to different thoughts expressed in the speech process, and the cognitive processes involved. In particular, pauses in the smallest syntactic units (e.g., between determiners and nouns, as in *the PAUSE bear*, or after auxiliaries, as in *It is PAUSE flowing out of the sink*) occur at a specific moment in mental time where a content word form or lexical concept has to be retrieved and inserted. At these moments, lexical and semantic memory has to be active (McDaniel et al., 2009), but there is virtually no syntactic structure to be built anymore: the relevant noun phrase (NP) or verb phrase (VP) has already been initiated and only the lexical head has to be inserted. At the initial boundary of an utterance or clause, by contrast, a complete syntactic configuration is about to be built that encodes a full thought. Such structures (except in high-frequency or idiomatic constructions remembered as wholes) are not memorized, but creatively produced on an occasion. In this way, the study of disfluencies at specific syntactic locations exploits a link between pausing and the thinking expressed in speech. It is of interest in dementia for this reason in particular.

In the context of AD, only two studies have focused on the syntactic context of pauses. Both studies investigated pause locations in populations in the earliest stages of AD. Each only examined two possible locations, and both found effects of syntactic position. Gayraud et al. (2011) found that early-stage AD participants paused more *within* clauses and phrases, but not at positions external to these minimal units of structure. On the other hand, Pistono et al. (2016) reported that a group with MCI due to AD produced more *between*-utterance pauses than healthy elderly controls. Both studies used an episodic recall task, though the former study involved long-term episodic memories, while the latter involved short-term ones. Notably, neither of these studies included the duration of pauses by syntactic position in their analyses. These results leave open the question of how pauses may manifest syntactically as the severity of dementia increases. Studies in other pathologies reinforce the motivation to study pauses at the level of syntax. Recent studies of aphasia (Zimmerer et al., 2019) and schizophrenia (Çokal et al., 2019a) have highlighted the embedded clause as a unit of special interest and detected comprehension deficits in embedded clause constructions. Çokal et al. (2019b) found that pausing patterns at clausal boundary positions distinguished schizophrenia groups with and without formal thought disorder from each other and from controls. A recent study evaluating language production in Huntington's dementia noted that disfluencies occurred largely along clausal boundaries rather than within-clauses, highlighting the importance of the clause as a unit of syntactic organization and an apparent problem in 'configuring thought-sized units' (Tovar et al., 2020, p.12; note the authors defined clauses in formal-linguistic terms as either complementizer phrases, CPs, or Tense Phrases, TPs).

The present study aimed to expand upon these findings by investigating both pause frequency and duration using a fine-grained annotation manual that subclassifies pauses by the forms of hierarchical syntactic complexity that they precede, moving gradually downwards from the largest such units (independent utterances) to embedded clauses, and finally to pauses within single clauses which we subcategorized for whether the subsequent lexical head was a noun, adjective or verb. This was motivated by the fact that such parts-of-speech distinctions, too, illuminate cognitive functions: the primary function of NPs is referencing objects, while VPs reference events and function grammatically as predicates, attributing properties to given referents (e.g., *[… [$_{VP}$ is falling of the stool]]*). Adjectives in turn function as predicates as well (whether sentential or adnominal). Gayraud et al. (2011) reported that early-stage AD participants paused much more before adjectives than healthy controls, citing the optionality of (attributive) adjectives as a potential explanation for this finding.

To further illuminate the cognitive function of pauses, we used VF scores at the single word level as a comparative measure, given that VF tasks specifically tap into lexical and semantic memory outside of a syntactic context. A picture description task was used to maximally isolate language from EM demands. To evaluate our measures against other studies and to investigate how fluency changes as dementia severity increases, this study includes a healthy elderly control, a mild AD, and a moderate AD group (outlined in methodology below). Our research questions were whether general fluency measures at the level of pauses and fillers, in both rate and duration, can distinguish groups; whether the syntactic positions in which they occur matter or not; and how such disfluencies at the level of spontaneous speech relate to VF scores. Our predictions were:

1. Pausing in spontaneous speech, including both *rate* and *duration*, will distinguish groups of healthy elderly controls from both mild and moderate AD and show sensitivity to disease progression as operationalized by Mini Mental State Exam (MMSE) scores.
2. Given attested SM decline in AD, altered pausing patterns in spontaneous speech will be particularly seen in the rate and/or duration of pauses occurring *within* clauses, when structure-building demands are lower than at the utterance and clausal levels and lexical retrieval demands (rather than utterance planning demands) are primary.
3. Based on evidence for a compensatory effect of pauses at early AD stages (Pistono et al., 2019), which manifested as a positive correlation between pause production and VF scores, we predicted that such a compensatory effect would be seen in mild AD. By contrast, a different pattern may be seen in moderate AD, given previous evidence of SM loss in AD. Specifically, a negative correlation between category VF scores and pause duration would indicate that as SM becomes more impaired, participants would produce less words in a category VF task and produce longer pauses in spontaneous speech as a reflection of difficulty in accessing semantic conceptual knowledge.

3. Methods

2.1 Participants

Samples were selected from the Pitt Corpus available on the Dementia Bank database (https://dementia.talkbank.org/), which provides audio recordings of AD participants and healthy controls gathered as part of the Alzheimer and Related Dementias Study at the University of Pittsburgh School of Medicine. The selected participants form three groups, 40 healthy elderly controls, 21 participants with mild to moderate AD, and 19 participants with moderate to severe AD as measured by the Mini-Mental State Exam (MMSE) (Folstein et al., 1975). The MMSE is a short test which evaluates cognitive function on a scale from 0-30, with a score of 30 indicating no cognitive impairment. Due to the ease of administration, the

MMSE is commonly used in clinical settings. Another popular neuropsychological test of cognitive function is the Clinical Dementia Rating (CDR). In contrast to the MMSE, the CDR includes measures related to daily functioning such as social interaction of home life. The CDR is expressed on a scale of 0-3, where 0 indicates no impairment and 3 indicates severe impairment (Morris, 1993, p. 2413).

All control participants had an MMSE score above 25 and a CDR score of 0. AD participants were selected across a range of MMSE scores from 8 to 23, divided into two groups, the cutoffs being 8-15 for 'moderate' and 16-24 for 'mild'. MMSE scores in the control group range from 26-30. Participants were selected that had no change to diagnosis status in subsequent visits, to avoid confounding diagnoses which contribute to dementia. Depression is a common comorbid condition with dementia. The Hamilton Depression Rating Scale (HAM-D) was administered at the time of the interviews, and these scores are available in the neuropsychological participant data of the Pitt Corpus. Because depression commonly occurs with dementia, mild depression scores were not used to exclude participants, and a HAM-D maximum threshold of 'moderate' (17+) was set for participant inclusion. We used the subscales used for the HAM-D proposed by Zimmerman et al. (2013). A Chi-Square goodness of fit test was performed to evaluate whether the proportion of participant sex was balanced across groups, and the differences were found to be insignificant $X^2(1, 80) = .450$, $p = .502$. Group differences were seen Age, and Education. Differences in age were expected because our groups are separated based on dementia severity, which increases over time. The age difference between the control and moderate group is the only age comparison which reaches statistical significance, and the effect size is moderate. Group differences in education are similar in significance and effect size for both the control/mild and the control/moderate comparisons. See Table 1 for group demographic and clinical data.

Table 1: Demographic and clinical data

	Control	Mild	Moderate	Control / Mild		Mild/Mod		Control / Mod	
		M (SD)		p	d	p	d	p	d
Age	67.5 (5.6)	70.1 (10.5)	70.8 (6.8)	.232	0.31	.957	0.02	.047	0.54
Ed.(yrs)	13.7 (2.2)	12.8 (2.6)	11.8 (3.3)	.033	0.57	.513	0.21	.037	0.57
MMSE	29.0 (1.1)	19.6 (2.2)	11.5 (1.8)	< .001	3.05	< .001	3.53	< .001	2.91
CDR	0.00 (0.0)	1.3 (.46)	1.9 (.32)	< .001	10.42	.513	1.56	< .001	16.59
HAM-D	1.9 (2.1)	7.4 (3.6)	6.7 (3.3)	< .001	1.94	.514	0.21	< .001	1.74

Note. The p-values presented are the result of Mann-Whitney U tests. The d-value presented is the Cohen's d measure of effect size.

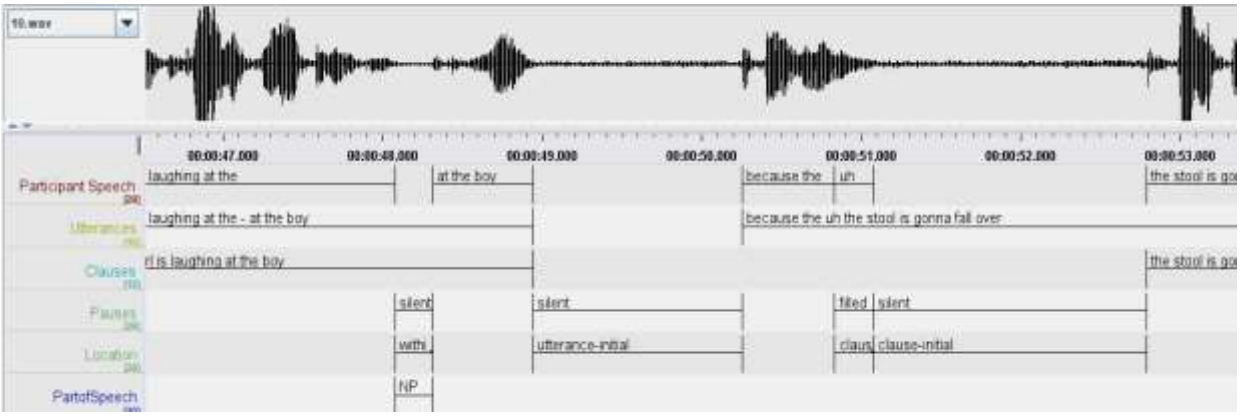
2.2 Procedure

Audio files of the Cookie Theft picture description task were downloaded for the selected participants from the DementiaBank in mp3 format, then converted to WAV and loaded into the ELAN (Brugman & Russel, 2004) software for manual annotation. Samples begin on the first turn-initial silence following instructions, showing that the participant has engaged in the task. In samples that begin with participant speech, the opening silence is trimmed, because there is no way to know how long the pause between instructions and speech is. Each sample terminates with the end of the final utterance by the participants, after which the investigator indicates the session is over. The final silence between the participant and investigator is excluded, as this is highly dependent upon the judgement of the investigator.

An ELAN template was designed to capture each variable to a separate tier, with all annotations time-aligned to the audio file. See Figure 1 below for a sample annotation using the template. Further examples can be found in the annotation manual included in supplementary materials. For each sample, the introductory segments pertaining to the communication of task expectations and concluding segments signaling the end of the task were excluded from analysis. The remaining, task-relevant segment was then transcribed orthographically. Prompts and responses from the investigator were transcribed separately from the participants' speech and excluded from the sample analysis. Pauses followed by investigator prompting have been marked as 'Prompt Initial Silence' and excluded from the count of total pauses, as the length of pauses before prompts is based on the investigator's judgement, and we cannot tell from an audio sample alone whether the silence necessarily constitutes disengagement. Additionally, there are instances throughout the samples in which the participant asks a question to the investigator. The silences in these interactions are more indicative of the linguistic behavior of the investigator than of the participant. All such silences are excluded, as the study seeks to identify patterns in participant disfluency.

All utterances and clauses were identified and tallied for each sample. For this study, utterances were defined as units of propositional information contributing new information to the discourse. Note that by this definition, utterances do not need to be clauses. Clauses were defined as a phrase consisting of a verb and minimally one nominal argument (dependent). All pauses and fillers of 200ms or longer were annotated.; A threshold of 200-250ms is typical minimum duration range for perceivable pauses (Zellner, 1994), and our choice is consistent with related studies (Gayraud et al., 2011; Pistono et al., 2019). Pauses longer than 3s, which are rare in neurotypical speech, were only annotated when they were not followed by an investigator prompt. All pauses were first marked for location and duration, then coded by manner: silent or filled. Fillers included the following: *hm*, *mm*, *uh*, and *um*.

Figure 1.



Example of ELAN annotation template and pause location marking

Note. Labels in the column on the left side of the figure are tier names within ELAN. Each value noted within a bracket is an annotation. This example illustrates the tagging of 3 silent pauses and one filler on the “Pauses” tier. Each of the three possible tags is seen here on the “Location” tier. The only pause marked on the “PartofSpeech” tier is that which was classified as “within-clause” on the “Location” tier

Scoring of the VF tests was performed following the manual detailed in Ledoux et al. (2014), which includes measures of clustering and switching. We chose this manual because it allows for the separation of cognitive functions involved in the VF tests. It is generally thought that clustering reflects the integrity of SM, whereas switching is related to executive function (Ledoux et al., 2014; Quaranta et al., 2019; Troyer et al., 1997). VF scores are provided in Table 2.

Table 2
Average rating of VF clinical assessments

		Mild	Mod	Mild / Mod	
Category VF		(n = 18)	(n = 15)	p	d
	Correct Words	10.17 (5.97)	5.07 (3.24)	.012*	1.26
	Mean Cluster	2.33 (1.52)	1.84 (1.72)	.087	0.77
	Switches	4.33 (4.12)	2.47 (2.80)	.324	0.49
Letter VF		(n = 19)	(n = 9)		
	Correct Words	7.11 (3.46)	3.11 (3.41)	.035*	0.99
	Mean Cluster	1.12 (1.77)	1.44 (1.33)	.478	0.34
	Switches	5.32 (3.77)	2.00 (2.74)	.073	0.82
Combined VF		(n = 18)	(n = 9)		

Correct Words	17.00 (8.44)	9.44 (4.90)	.055	0.88
Mean Cluster	1.92 (0.84)	1.88 (1.60)	.361	0.43
Switches	9.56 (5.95)	5.44 (4.19)	.186	0.58

Note. Reported p-values are the result of a Mann-Whitney U test after a Benjamini-Hochberg correction. The rightmost column is Cohen's d effect size measure.

2.3 Measures

The variables measured can be divided into three groupings of increasingly fine-grained detail. The first set of variables are general fluency measures, including speech, pause, and fillers. Participant speech was divided into utterances, and then clauses, and the count of utterances and clauses were used to control for sample length differences. All pauses and fillers were time-aligned and measured for duration Further, each silent pause was categorized by the syntactic context in which it occurred. The syntactic positions identified were utterance-initial, clause-initial, and within-clause. All three of these positions can be observed in the example in Fig. 1 above. These locations represent a hierarchy of complexity, with within-clause being lowest in the hierarchy, as they are located within the smallest grammatical unit considered here (a single verbal, prepositional, or nominal phrase). Such phrases are proper parts of clauses, which in turn are proper parts of utterances. Utterance-initial pauses were essentially 'between-utterances', as turn-initial pauses (silence following an investigator prompt) were excluded. It is worth noting that when participants produced fragmented speech (incomplete utterances, either abandoned or reformulated), the fragments were not annotated as utterances. However, pauses before such false starts and truncations were marked utterance initial. Due to this, the rate of utterance-initial pauses increased when a participant produced more fragmented speech.

The final grouping of variables subcategorized within-clause pauses based on the lexical category of the content word following the pause (N, V, or A, for nouns, verbs, and adjectives/adverbs). An independent rater individually annotated 10% of the audio samples for utterances, clauses, pause location, and part of speech tags. After review, there was an inter-rater point-by-point agreement of 87.5% for both utterances and clauses individually. For reliability of pause location and part of speech tags, which are categorical distinctions, Cohen's kappa was used to account for chance agreement. Pause location and part of speech showed high degrees of concordance between raters (κ = 90.3) and (κ = 94.3), respectively. See Table 3 for a summary of measures and their calculation methods.

Table 3

Fluency variables and calculation methods

Variable	Calculation Method
General Fluency	
Mean length of utterance (words)	Word count / utterance count
Words per minute	Word count / (total utterance duration /1000) /60)
Pause rate	Pause count / utterance count
Filler rate	Filler count / utterance count
Pause duration	Total pause duration (ms) / pause count
Filler duration	Total filler duration (ms) / filler count
Syntactic Location of Pauses	
Utterance-initial (UI) rate	UI count / utterance count
Clause-initial (CI) rate	CI count / clause count
Within-clause (WC) rate	WC count / clause count
Utterance-initial (UI) duration	Total UI duration (ms) / UI count
Clause-initial (CI) duration	Total CI duration (ms) / CI count
Within-clause (WC) duration	Total WC duration (ms) / WC count
Within-clause Pauses	
N-initial (NI) rate	NI count/ clause count
V-initial (VI) rate	VI count/ clause count
A-initial (AI) rate	AI count/ clause count
N-initial (NI) duration	Total NI duration (ms) / NI count
V-initial (VI) duration	Total VI duration (ms) / VI count
A-initial (AI) duration	Total AI duration (ms) / AI count

2.4 Data Analyses:

Descriptive statistical analysis was performed to establish group averages. Mann-Whitney U tests were selected for between-group comparisons, due to the non-normal distribution of the data. Cohen's d was used to assess effect sizes along with the Mann-Whitney U tests. Pairwise comparisons were run between all groups. Mahalanobis distances were used to detect multi-variate outliers for the fluency metrics. Excluding the 3 cases which were flagged as outliers did not change the pattern of significant results, and we therefore chose not to exclude these cases. The analysis of VF scores was performed on the combined test scores and the category and letter subtest scores. Mann-Whitney U tests were run to compare the mild and moderate groups within the subset of AD participants where VF tests where available. To correct for multiple comparisons, we used the Benjamini-Hochberg (1995) method, by multiplying the p-values by the total number of comparisons divided by the p-value rank. A Spearman's correlation was performed for the mild and moderate groups separately. No correction was applied separately for this correlational analysis as it was based on an a priori hypothesis. A syntactic analysis of filler locations was not included due to a poverty of data. Fillers were not produced by all participants, and the filler rate was low compared to the pause rate.

4. Results

Descriptive statistics, p-values resulting from group comparisons using the Mann-Whitney U test, and Cohen's d for effect sizes can be found in Table 4; see also Figures 2-5. As expected, at the level of general fluency measures, both AD groups produced a significantly higher pause rate and lower speech rate, as measured in words per minute (WPM), in comparison to the control group, with large effect sizes of 0.8 (Cohen's d) or higher. On the other hand, no significant differences were found for pause *duration*, or for either filler rate or duration. The analysis of pauses by syntactic locations revealed significant group differences across utterance-initial, clause-initial, and within-clause positions, hence boundaries of units at all levels of structural complexity, which at the highest level (utterances) enclose full thoughts. Largest effect sizes were seen in pause rate in this utterance-initial position, where controls differed from both AD groups. Measures of rate and duration of pauses again patterned differently in this respect, as differences between controls and either AD group in the duration of utterance-initial pauses were completely non-significant.

Moving down in the levels of syntactic complexity to utterance-internal positions, controls and moderate-AD differed in clause-initial but not within-clause positions, consistently in both rate and duration, while controls and mild-AD showed the reverse pattern: they differed in within-clause but not clause-initial positions, again in both rate and duration. In line with this pattern, when finally zooming into the within-clause positions in a more fine-grained way, all differences were confined to the comparison of controls and mild-AD, and again strikingly, to the N-initial position only, with strong effect sizes (p = .017, d = .885 and p = .016, d = .874, for rate and duration of NI-pauses, respectively).

Table 4

Fluency Measure	Descriptive Stats by Group as M(SD)			Mann-Whitney U *p*-values and Cohen's *d* effect size					
	Control (n = 40)	Mild (n = 21)	Moderate (n = 19)	Control / Mild		Mild / Mod		Control / Mod	
		M (SD)		*p*	*d*	*p*	*d*	*p*	*d*
General fluency									
Sample duration (ms)	56566 (26002)	65868 (46422)	86340 (50986)	.797	.094	.337	.449	.105	.586
Total word count	102.70 (52.27)	83.19 (48.99)	92.52 (55.94)	.150	.523	.701	.183	.555	.225
MLU	7.74 (1.50)	7.8 (1.34)	7.49 (1.73)	.889	.053	.566	.276	.572	.227
WPM	114.02 (31.03)	85.9 (30.68)	89.86 (39.59)	.017*	.852	.978	.009	.031*	.793
Pause rate	1.17 (.39)	1.7 (.63)	1.79 (.75)	.005*	1.081	.839	.096	.006*	1.122
Filler rate	0.19 (.16)	0.3 (.38)	0.23 (.33)	.653	.178	.587	.253	.734	.114
Pause duration	1756.52 (914.94)	2007 (1044.42)	1759.13 (886.10)	.496	.265	.564	.267	.936	.026
Filler duration	375.11 (242.06)	296.23 (211.23)	247.15 (213.21)	.261	.404	.685	.194	.155	.513
Syntactic location of pauses									
UI rate	0.75 (.19)	1.01 (.28)	1.11 (.32)	.005*	1.033	.458	.357	.002*	1.285
CI rate	0.03 (.06)	0.06 (.06)	0.08 (.08)	.244	.418	.567	.290	.037*	.751
WC rate	0.23 (.20)	0.42 (.26)	0.39 (.32)	.015*	.866	.689	.183	.089	.631
UI duration	2129.43 (1245.9)	2572.46 (1763.5)	2146.64 (1132.2)	.451	.293	.693	.196	.938	.034
CI duration	230.45 (385.76)	805.6 (1409.18)	546.61 (464.56)	.158	.488	.899	.067	.039*	.753
WC duration	554.09 (319.13)	908.07 (365.08)	794.56 (464.03)	.004*	1.075	.442	.375	.152	.511
Within-clause pauses									
NI rate	0.1 (.10)	0.25 (.18)	0.21 (.19)	.017*	.885	.647	.219	.089	.625
VI rate	0.09 (.11)	0.16 (.16)	0.15 (.15)	.159	.511	.940	.035	.159	.515
AI rate	0.03 (.05)	0.01 (.03)	0.03 (.05)	.156	.524	.349	.437	.680	.145
NI duration	464.1 (396.18)	934.61 (610.19)	606.4 (438.91)	.016*	.874	.160	.618	.517	.258
VI duration	439.1 (415.78)	742.19 (696.49)	805.33 (670.19)	.177	.469	.686	.180	.108	.589
AI duration	226.69 (361.09)	170.73 (416.62)	284.58 (565.69)	.373	.332	.560	.288	.789	.102

Group results and between-group comparisons using Mann-Whitney U tests

Note. MLU: Mean length of utterance (words); WPM: Words per minute; UI: Utterance-initial; CI: Clause-initial; WC: Within-clause; NI: Noun-initial; VI: Verb-initial; AI: Adjective/Adverb initial.

Figure 2

Pause rates by syntactic location

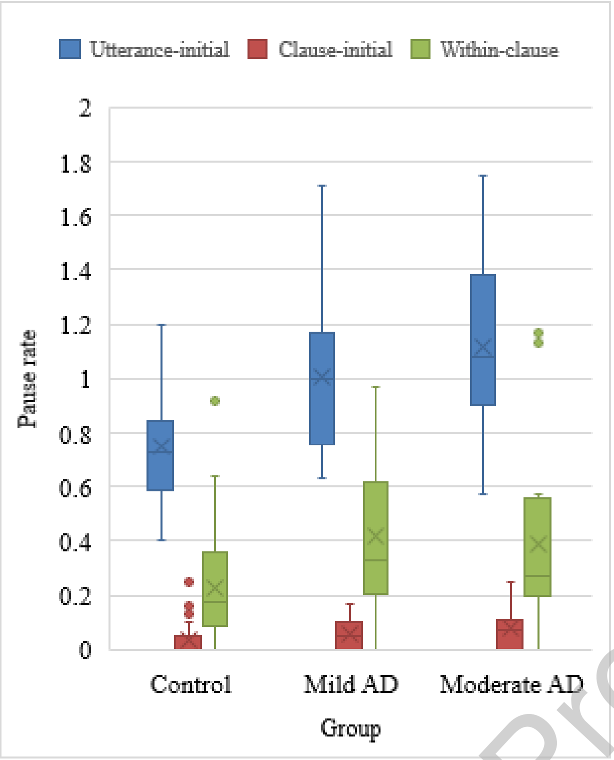


Figure 3

Pause duration by syntactic location

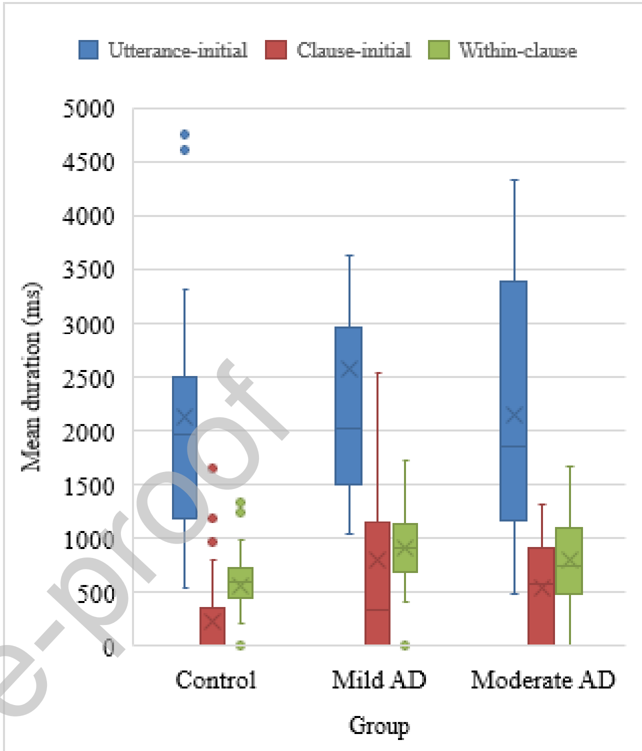


Figure 4

Within-clause pause rates by part of speech

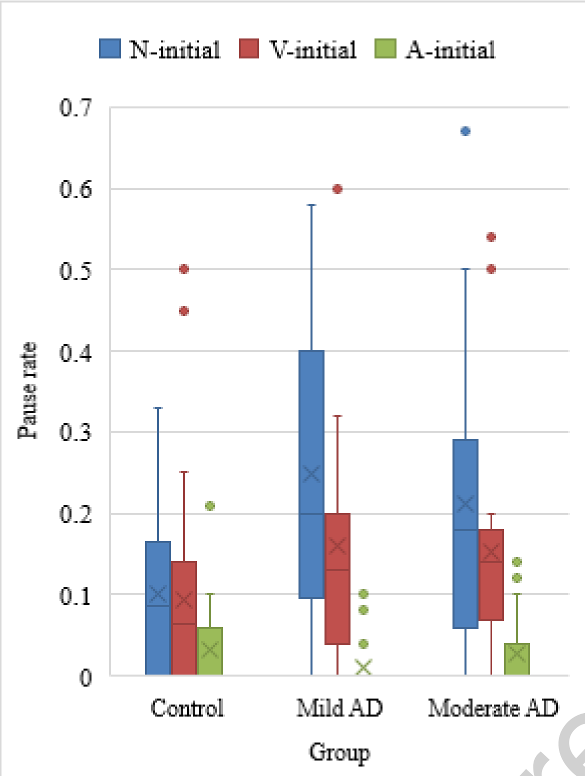
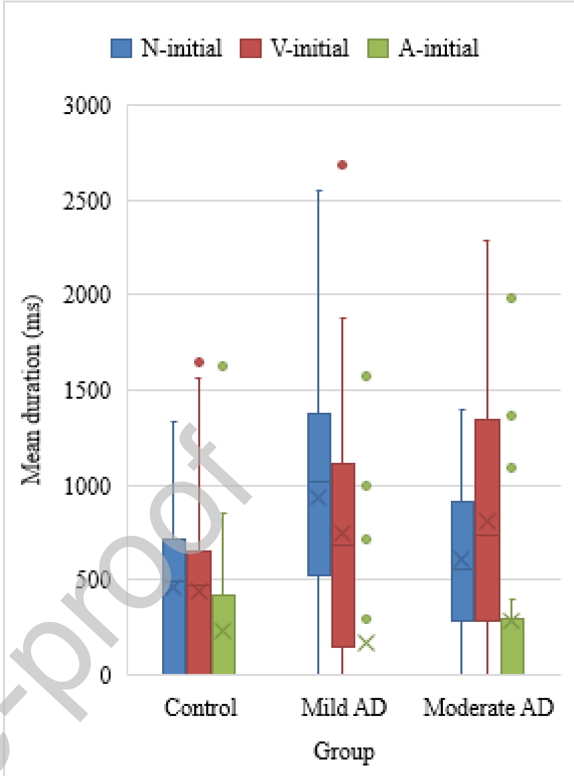


Figure 5

Within-clause pause duration by part of speech



The Spearman's correlations run within the AD groups revealed only one correlation which retained significance after controlling for education and Hamilton depression scores. This correlation was between noun-initial pause duration and MMSE scores in the moderate group, $r_s(19) = .66$, $p = .004$.

Spearman's correlations between category VF scores at the single-word level and pause rate measured in connected speech revealed significant correlations in the mild-AD group, but none in the moderate-AD group. Within the mild-AD group, pause rate was negatively correlated with the number of correct words produced in both category and letter VF tasks (as well as the combined VF measures). For the mild-AD group, pause rate was also negatively correlated to the mean cluster measure on the letter VF task and switch count in the combined VF scores. A correlation matrix highlighting significant relationships can be found in Table 5.

Table 5:

Spearman's correlation matrix for pause rate and VF scores by group

		Category VF			Letter VF			Combined VF		
		Word	Cluster	Switch	Word	Cluster	Switch	Word	Cluster	Switch
Mild	R_s	-.504	-.096	-.409	-.507	-.632	-.450	-.565	-.286	-.663
	p =	.033*	.704	.092	.027*	.004*	.053	.015*	.250	.003*
Mod	R_s	-.022	-.270	.015	-.237	-.358	.122	.059	-.297	.167
	p =	.939	.330	.958	.539	.343	.754	.881	.437	.668

Note. For full correlational matrices see Supplementary Materials: S2 (category VF), S3 (letter VF), and S4 (combined VF).

5. Discussion

This study sought to enhance current insights on the disfluency profile in AD extracted from picture descriptions of people with either mild or moderate AD, by identifying syntactic locations of disfluencies in spontaneous speech and linking these to measures of verbal fluency (VF) at the single word level. Results partially confirmed our first broad prediction that groups would be distinguishable based on measures of pauses and fillers at a general level of analysis (i.e., not considering their specific syntactic positions). At the same time, they revealed striking differences in the patterning of rates and durations, since overall pause *rate* but not *duration* differed across groups. This finding is not consistent with the results from Pistono et al. (2019), in which both pause rate and duration distinguished groups on a picture-based narrative. However, it is generally consistent with previous studies (Gayraud et al., 2011; Pistono et al., 2016; 2019) insofar as rate and duration need not pattern the same, although the divergence between overall pause rate and duration has not patterned consistently across studies. In Gayraud et al. (2011), it was the total number of pauses and not pause duration which differed significantly between early AD and healthy control groups. In the Pistono et al. (2016; 2019) studies, it was pause duration, and not rate which differentiated MCI and early AD groups from healthy controls. These contradictory patterns came from a pause analysis of speech produced in memory-based narratives. A notable methodological difference may be that in the Pistono et al. (2016, 2019) studies, pauses and fillers were counted together, while in Gayraud et al. (2011), like here, they were counted separately. In the present study, as in Gayraud et al. (2011), pauses and fillers parted ways, with the latter showing *no* significant differences across groups for either rate or duration. This pattern reveals that when the flow of speech is broken through a disfluency in AD, this tends to happen specifically through an *empty* pause: no effort is made to bridge such 'gaps', of a kind that would be revealed by an increase of fillers. Fillers have been functionally related to speech monitoring (Levelt, 1983) and hearer-oriented social signaling (Lake et al., 2011), particularly in utterance-initial positions (Howes et al., 2017), suggesting that in AD, these processes do not increase as and when interruptions of the speech generation process start to appear more often. Put differently, the gaps opening are 'true gaps', in the sense that they are not monitored or metacognitively processed. It has also been proposed that part of the hearer-oriented social signaling conveyed by fillers can be intended to allow extra time for lexical retrieval (Clark & Fox, 2002). Our results thus indicate that if lexical retrieval difficulties are experienced by AD participants, such an increased difficulty is not reflected by the use of fillers.

The differences found here between general pause rate and duration of pauses are harder to interpret based on evidence available at present. These differences were not only observed when considering overall pause rate and duration, but also when considering pauses subcategorized as occurring in utterance-initial positions: group comparisons were significant in relation to utterance-initial pause rate but not duration (Table 4). The crucial question is which cognitive mechanisms rate and duration index. Conceptually, the rate of pauses indicates how often a problem occurs, while the duration of a pause indicates how long it takes to solve it. However, this conceptual difference does not preclude possible consistencies between them, as were indeed found in clause-initial and within-clause positions. Moreover, rate was calculated by pause count divided by utterance count, and average duration as total pause duration divided by pause count. Hence both measures were derived from a shared data point. This decision made sense because two relevant questions are (i) how often pauses occur given how many utterances are produced, and (ii) how long they last given how often they occur. Despite the shared data point, the fact remains that they patterned differently in crucial respects. Our tentative suggestion is that no duration effect shows for the overall pause measures, because no such duration effect shows for utterance-initial pauses (which were the most frequent).

Future studies should thus specifically focus on this position. There are factors involved in the latter that were not controlled for in our annotation scheme and could illuminate the cognitive mechanisms involved. In particular, the utterance-initial position was a relatively broad syntactic measure, consistent with the fact that utterances can be affected using a wide range of possible syntactic phrase types. In addition, utterance-initial pauses included those before disfluent speech, as noted (i.e., false starts, repetitions, truncations). Due to the high significance and effect size seen for the utterance-initial rate measure, it would be worthwhile to implement a more fine-grained analysis in future studies, which teases out the potential differences of pause function in this position. Specifically, pauses preceding disfluent speech such as repetitions, false starts, and truncations could be separated from pauses preceding fully fluent utterances. Such a distinction would enhance interpretability, as a pause preceding an independent clause may be interpretable as a planning pause, whereas pauses before repetitions likely represent a distinct phenomenon.

Our second broad prediction concerned group differences with regards to pausing in different syntactic positions, with a specific prediction for within-clause positions given noted SM impairments in AD. In contrast to this prediction, we found that the utterance-initial pause rate was increased in both AD groups relative to controls, thus showing a consistent result across the AD groups. Further, the utterance-initial rate increased from the mild to moderate stage, and while this difference was not significant when directly comparing the AD groups, the significance and effect size in comparison to controls did increase with severity. The rate of clause-initial pauses also increased from the mild to moderate phase, although this measure does not significantly distinguish AD participants from controls until the moderate phase. This pattern seems to indicate that an increase in clause-initial pause production begins in the early stage of AD and progresses at a slow rate, rather than suddenly appearing at the moderate stage of AD: if the problem only began at the moderate stage, we would expect to see a between-group difference in the comparison of mild and moderate AD groups.

In the within-clause pause position, on the other hand, we observed a distinctive break from the pattern of pause rate increasing with severity observed in the utterance-initial and clause-initial measures. Thus, the rate and duration of pauses in the within-clause position peaked in the mild AD group and then decreased again in the moderate group. While the between-AD-group comparison did again not differentiate mild from moderate AD on these measures, relations between these groups and controls signal differences between them: the within-clause measures distinguished mild AD from controls, but no longer did so for the moderate AD group. We are unable to compare these patterns to prior studies (Gayraud et al. 2011, Pistono et al. 2016), as the duration of pauses by syntactic position has not previously been reported, nor did these studies include more than one participant group.

Overall, our conclusions regarding the syntactic positions of pauses are that an increase of pausing in utterance-initial and clause-initial positions in AD lends no support to the idea that effects are mainly driven by problems of lexical retrieval: They concern structure-building at a grammatical level. Specifically, what utterances as defined here have in common is that they express a complete thought, regardless of whether or not such a thought is syntactically expressed as a clausal configuration. Given our results, problems in configuring the syntactic unit that corresponds to a complete thought might be said to be more of a ‘trait’-marker of cognitive decline in AD, while pauses in within-clause positions may be more transitory and seen at the early stages. Clause-initial positions are similar to utterance-initial positions in that neither utterance- nor clause-sized syntactic units are (typically) memorized. Rather, at these junctions, complex structural configurations are being put together creatively, as sense is being made of the objects seen, of how they relate within complex events of which the objects form parts, and how the events in turn relate to each other in complex temporal and causal relations (e.g., the boy *falling of the stool,* while the water is *running out of the sink*). Pauses at boundaries preceding utterances have been associated with content planning (Corley et al., 2007; Goldman-Eisler, 1958; Levelt, 1989), and the occurrence of pauses in these positions has also been shown to increase along with syntactic complexity (Ferreira, 1991) . Altered pausing in these positions therefore is suggestive of problems in a process of structural assembly of complex event representations. In particular, no EM demands are involved, as the events are visually observed in the scene. Put differently, the pattern is clearly not that of people merely not remembering what objects are called, delaying the production of a given content word within an already initiated phrase. In short, this pattern of pausing in utterance- and clause-level position does not support the idea of a secondary effect of SM or EM impairment.

Our third prediction concerned correlations between disfluency measures in spontaneous speech and VF at the single-word level. Correlations of this type cross between two very different linguistic levels, the former involving structure-building at a grammatical level, the latter confined to lexical retrieval at a semantic (category fluency) or phonetic (letter fluency) level. The correlational patterns we found in this respect were confined to the mild group only, and were obtained for both category and letter fluency, suggesting no problem specific to SM, as measured by category fluency, or general to the AD. As such, this pattern arguably further corroborates the above conclusion about semantic or lexical memory deficits not being the primary drivers of temporal disfluencies. As noted in the introduction, Pistono et al. (2019) did find a correlation between semantic fluency test scores at the single word level and pause rate in spontaneous speech, however in an MCI group. This partially motivated our study design. However, the correlation in that previous study was positive, supporting the idea that increased pause rate reflects a compensation effect, while our observed correlations were in the opposite direction: silent pause rate correlated *negatively* to VF score in the mild AD group. This finding fails to support our final prediction based on Pistono et al.'s (2019) compensatory mechanism account of pausing in early AD.

A further reason against the conclusion that a primary SM impairment explains pausing patterns, even in the mild-AD, is that the within-clause pausing pattern was specifically driven by a noun-effect specific to this early-stage group. N-initial rate and duration measures distinguish the control from mild-AD with a strong effect size ($p < .000$, $d = .89$ and $p < .000$, $d = .87$, respectively), with no significant differences in either V-initial or A-initial positions. Evidence for an early-stage noun-effect is strengthened by the positive correlation between noun-initial pause duration and MMSE scores in the moderate AD group. This result indicates that after individuals move from the mild to the moderate stage of AD, the pauses before nouns begin to shorten in duration as dementia severity increases. We speculate that this shortening and gradual disappearance of pauses reflect the receding of an initial cognitive effort devoted to solving an emerging processing problem at this local level of structural complexity, which is indexed by such pauses. In other words, the effort is abandoned as the disease progresses, with later stages revealing cognitive more in the syntactically larger units, such as clauses and utterances.

Since the effect in question is a parts-of-speech effect, it entails that the problem is grammatically specific rather than a generalized problem of lexical or semantic memory – V- and A-positions involve content words to be remembered as much as N-positions do. On the other hand, the effect is consistent with a more specialized lexical or semantic memory problem, which would only affect object concepts since nouns prototypically denote objects. However, it is not easy to differentiate our notions of what is semantically an 'object' from that of 'what is denoted by a noun', and indeed our annotation scheme was focused on nouns, which does not coincide with the notion of denoting an object. Where N-positions differ is in how they function. As noted in the introduction, the primary function of nouns in grammar is to provide a lexical basis for referentiality, while verbs and adjectives/adverbs primarily function as predicates. The possibility of a primary problem with referentiality to objects in AD is thus an important new question for research on language in AD: there has been independent evidence of problems with referentiality, as when participants produce overly vague NPs (e.g., *thing*, *stuff*) in the place of semantically more specific nouns, or replace lexical nouns by pronouns (e.g., Ostrand & Gunstad, 2020; Chapin et al., 2021).

The present findings differ from that of two previous studies of the syntactic locations of pauses in connected speech in AD. Recall that Gayraud et al. (2011) found more pauses within clauses and phrases in early AD participants but reported no significant result for pauses external to these minimal units. On the other hand, Pistono et al. (2016) reported that a group with MCI due to AD produced more between-utterance pauses than healthy elderly controls. These studies are not directly comparable to the present one, for several reasons. Firstly, the Pistono et al (2016) study did not separate silent pauses from fillers, as we have done in the present study; however, this distinction was made in the measures of Gayraud et al (2011). Secondly, whereas we calculated our pause rates by utterance, Pistono et al. (2016) calculated pause rates per 100 words and Gayraud et al. (2011) reported pauses measures as percentages rather than rates. Finally, apart from their sample sizes being smaller, both studies rated connected speech of early AD participants in EM tasks, and our participants had more severe dementia scores. While Pistono et al. (2016) analyzed a group with an average MMSE score of 24.5 ± 2.9 and the participants in Gayraud et al. (2011) had a mean MMSE of 22.6 ± 2.5, our mild AD group has a much lower mean MMSE of 19.6 ± 2.2 and our moderate group had a mean MMSE of 11.5 ± 1.8. One reason for the discrepancy we found is therefore likely a trajectory of linguistic decline observed in AD. In the present study, both the rate and duration of within-clause pauses increased in the mild-AD group relative to the control group, but these numbers dropped back down with severity, and no longer differentiated the moderate group from controls, as noted above. This pattern is suggestive of the idea that early AD is different, specifically in how increased pausing reveals the cognitive effort of grappling with an emerging disease. In more advanced stages, this effect disappears from lower syntactic levels that are more lexical, shifting to the higher syntactic level of clauses and utterances, indicating a potential problem in structure-building.

Limitations and future directions

One limitation of our study was the relatively short length of the audio samples, which relates to the nature of the task. An additional limitation is the exclusion of disfluent speech from the analyses in the present study, which could further illuminate cognitive mechanisms involved. It would be specifically interesting to include repair disfluencies in the context of pause measures.

The general lack of statistical differences between the mild and moderate AD groups has important implications for future work. Firstly, it is possible that the measures used in the present study simply lacked the sensitivity required to distinguish these groups. Future work could thus benefit from more fine-grained measures of the syntactic context of pauses.

Another possible implication is that the onset of the problems indexed by the syntactic context of pauses may precede the mild AD phase, motivating further investigation into these measures in prodromal stages of dementia, such as subjective cognitive complaints (SCC) and MCI.

It is finally worth noting that depression scores present an important confound in this clinical population. As stated in our first prediction, we expected that overall pause rate and duration would increase with dementia severity. However, none of the general pause or filler measures remained significant after controlling for depression scores. Depression is a confound difficult to avoid when researching dementia groups, particularly beyond the mild stage, yet should be kept in mind in any future studies on the subject. Regarding the early-noun effect observed in our data, further investigation is required to determine the influence of other potential confounds, such as lexical frequency.

Conclusions

In sum, the present study confirms highly significant differences in pausing behaviour, with empty pauses opening across syntactic positions and likely not due to a specific effect of memory, whether episodic or semantic. Differential pause effects found across different syntactic positions and measures not only cause explanatory problems for an SM or EM-based account, but also for accounts based on generalized PS (Salthouse, 1996), WM (Salthouse, 1996; Martin & Slevc, 2014), or visuo-spatial (Binetti et al., 1998) cognitive capacities. We also found no clear support of prior evidence (Pistono et al., 2019) of a compensatory mechanism visible in pausing behaviour, though a difference in within-clause pausing patterns in mild but not moderate AD clearly suggests that in early disease stages, pauses play a different cognitive role than in later ones, and lexical retrieval mechanisms may be among the factors involved.

CRediT author statement

Mary Lofgren: Conceptualization, Methodology, Investigation, Formal analysis, visualization, Writing- Original draft, Writing- Review and editing, Funding acquisition. **Wolfram Hinzen**: Conceptualization, Methodology, Supervision, Writing- Original draft, Writing- Review and editing, Funding acquisition.

6. References

Ahmed, S., Haigh, A. M., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain : a journal of neurology*, *136*(Pt 12), 3727–3737. https://doi.org/10.1093/brain/awt269

Ash, S., Evans, E., O'Shea, J., Powers, J., Boller, A., Weinberg, D., Haley, J., McMillan, C., Irwin, D. J., Rascovsky, K., & Grossman, M. (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology*, *81*(4), 329–336. https://doi.org/10.1212/WNL.0b013e31829c5d0e

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995; 57(1):289±300.

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected Speech in Neurodegenerative Language Disorders: A Review. *Frontiers in psychology*, *8*, 269. https://doi.org/10.3389/fpsyg.2017.00269

Brugman, H., & Russel, A. (2004). *Annotating Multi-media / Multi-modal resources with ELAN*. Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation. https://archive.mpi.nl/tla/elan

Butterworth, B. (1979). Hesitation and the Production of Verbal Paraphasias and Neologisms in Jargon Aphasia. *Brain and Language*, *8*, 133–161.

Chapin, K., Clarke, N., P. Garrard, P., & Hinzen, W. (2021). A more fine-grained linguistic profile of Alzheimer's disease and Mild Cognitive Impairment. *Under review.*

Çokal, D., Zimmerer, V., Varley, R., Watson, S., & Hinzen, W. (2019a). Comprehension of embedded clauses in schizophrenia with and without Formal thought disorder. *The Journal of nervous and mental disease*, *207*(5), 384–392. https://doi.org/10.1097/NMD.0000000000000981

Çokal, D., Zimmerer, V., Turkington, D., Ferrier, N., Varley, R., Watson, S., & Hinzen, W. (2019b). Disturbing the rhythm of thought: Speech pausing patterns in schizophrenia, with and without formal thought disorder. *PloS one*, *14*(5), e0217404.

DeDe, G., & Salis, C. (2020). Temporal and episodic analyses of the story of Cinderella in latent aphasia. American Journal of Speech-Language Pathology, 29(1S), 449-462. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0210

de Lira, J. O., Ortiz, K. Z., Campanha, A. C., Bertolucci, P. H., & Minett, T. S. (2011). Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *International psychogeriatrics*, *23*(3), 404–412. https://doi.org/10.1017/S1041610210001092

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in human neuroscience, 6*, 69. https://doi.org/10.3389/fnhum.2012.00069

Emery V. O. (2000). Language impairment in dementia of the Alzheimer type: a hierarchical decline?. *International journal of psychiatry in medicine, 30*(2), 145–164. https://doi.org/10.2190/X09P-N7AU-UCHA-VW08

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*(2), 210–233. https://doi.org/10.1016/0749-596X(91)90004-4

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State" A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry*, *12*, 189-198.

Garrard P, Maloney LM, Hodges JR, Patterson K. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. Brain. 2005 Feb;128(Pt 2):250-60. https://doi.org/10.1093/brain/awh341

Gayraud, F., Lee, H. R., & Barkat-Defradas, M. (2011). Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. *Clinical linguistics & phonetics*, *25*(3), 198–209. https://doi.org/10.3109/02699206.2010.521612

Goldman-Eisler, F. (1958). Speech Production and the Predictability of Words in Context. *Quarterly Journal of Experimental Psychology*, *10*(2), 96–106. https://doi.org/10.1080/17470215808416261

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia*, *42*(9), 1212–1222. https://doi.org/10.1016/j.neuropsychologia.2004.02.001

Howes, C., Lavelle, M., Healey, P., Hough, J., & McCabe, R. (2017). Disfluencies in dialogues with patients with schizophrenia. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society, 26-29 July 2017* (pp. 2230-2235)

Joubert S, Brambati SM, Ansado J, Barbeau EJ, Felician O, Didic M, Lacombe J, Goldstein R, Chayer C, Kergoat MJ (2010) The cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early Alzheimer's disease. Neuropsychologia 48, 978-988. https://doi.org/10.1016/j.neuropsychologia.2009.11.019

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic bulletin & review*, *18*(1), 135–140. https://doi.org/10.3758/s13423-010-0037-x

Lang, C. J., & Quitz, A. (2012). Verbal and nonverbal memory impairment in aphasia. *Journal of neurology*, *259*(8), 1655–1661. https://doi.org/10.1007/s00415-011-6394-1

Laske, C., Sohrabi, H. R., Frost, S. M., López-de-Ipiña, K., Garrard, P., Buscema, M., Dauwels, J., Soekadar, S. R., Mueller, S., Linnemann, C., Bridenbaugh, S. A., Kanagasingam, Y., Martins, R. N., & O'Bryant, S. E. (2015). Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *11*(5), 561–578. https://doi.org/10.1016/j.jalz.2014.06.004

Ledoux, K., Vannorsdall, T. D., Pickett, E. J., Bosley, L. V., Gordon, B., & Schretlen, D. J. (2014). Capturing additional information about the organization of entries in the lexicon from verbal fluency productions. *Journal of clinical and experimental neuropsychology*, *36*(2), 205–220. https://doi.org/10.1080/13803395.2013.878689

Levelt W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press.

Liu, J., Zhao, J., & Bai, X. (2020). Syntactic Impairments of Chinese Alzheimer's Disease Patients from a Language Dependency Network Perspective. *Journal of Quantitative Linguistics*. 28(3), 253-281. https://doi.org/10.1080/09296174.2019.1703485

López-de-Ipiña, K., Alonso, J. B., Travieso, C. M., Solé-Casals, J., Egiraun, H., Faundez-Zanuy, M., Ezeiza, A., Barroso, N., Ecay-Torres, M., Martinez-Lage, P., & Martinez de Lizardui, U. (2013). On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. *Sensors*, *13*(5), 6730–6745. https://doi.org/10.3390/s130506730

Mack, J. E., Chandler, S. D., Meltzer-Asscher, A., Rogalski, E., Weintraub, S., Mesulam, M. M., & Thompson, C. K. (2015). What do pauses in narrative production reveal about the nature of word retrieval deficits in PPA?. *Neuropsychologia*, *77*, 211–222. https://doi.org/10.1016/j.neuropsychologia.2015.08.019

Martin, R. C., & Slevc, L. R. (2014). Language production and working memory. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), The Oxford handbook of language production (p. 437-450). Oxford University Press.

Martínez-Sánchez, F., Meilán, J. J., García-Sevilla, J., Carro, J., & Arana, J. M. (2013). Oral reading fluency analysis in patients with Alzheimer disease and asymptomatic control subjects. *Neurologia*, *28*(6), 325–331. https://doi.org/10.1016/j.nrl.2012.07.012

McDaniel, D., McKee, C., & Garrett, M. F. (2010). Children's sentence planning: syntactic correlates of fluency variations. *Journal of child language*, *37*(1), 59–94. https://doi.org/10.1017/S0305000909009507

McDowd, J., Hoffman, L., Rozek, E., Lyons, K. E., Pahwa, R., Burns, J., & Kemper, S. (2011). Understanding verbal fluency in healthy aging, Alzheimer's disease, and Parkinson's disease. Neuropsychology, 25, 210-225. https://doi.apa.org/doi/10.1037/a0021531

Mertens, P. (2004). The Prosogram: Semi-Automatic Transcription of Prosody Based on a Tonal Perception Model. *Proceedings of Speech Prosody*, (pp. 23-26).

Mirdamadi, F. S., & De Jong, N. H. (2015). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research, 31*(1), 105-116. https://doi.org/10.1177/0267658314554498

Morris, J. C. (1993, November). The Clinical Dementia Rating (CDR): Current version and scoring rules. Neurology, 43, 2412-2414. https://doi.org/10.1212/WNL.43.11.2412-a

Nilakantan, A. S., Voss, J. L., Weintraub, S., Mesulam, M. M., & Rogalski, E. J. (2017). Selective verbal recognition memory impairments are associated with atrophy of the language network in non-semantic variants of primary progressive aphasia. *Neuropsychologia*, *100*, 10–17. https://doi.org/10.1016/j.neuropsychologia.2017.04.006

Orimaye, S. O., Wong, J. S., Golden, K. J., Wong, C. P., & Soyiri, I. N. (2017). Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC bioinformatics*, *18*(1), 18-34. https://doi.org/10.1186/s12859-016-1456-0

Ostrand, R., & Gunstad, J. (2020). Using Automatic Assessment of Speech Production to Predict Current and Future Cognitive Function in Older Adults. *Journal of geriatric psychiatry and neurology*, 891988720933358. Advance online publication. https://doi.org/10.1177/0891988720933358

Pistono, A., Jucla, M., Barbeau, E. J., Saint-Aubert, L., Lemesle, B., Calvet, B., Köpke, B., Puel, M., & Pariente, J. (2016). Pauses During Autobiographical Discourse Reflect Episodic Memory Processes in Early Alzheimer's Disease. *Journal of Alzheimer's disease : JAD*, *50*(3), 687–698. https://doi.org/10.3233/JAD-150408

Pistono, A., Pariente, J., Bézy, C., Lemesle, B., Le Men, J., & Jucla, M. (2019). What happens when nothing happens? An investigation of pauses as a compensatory mechanism in early Alzheimer's disease. *Neuropsychologia*, *124*, 133–143. https://doi.org/10.1016/j.neuropsychologia.2018.12.018

Quaranta, D., Piccininni, C., Caprara, A., Malandrino, A., Gainotti, G., & Marra, C. (2019). Semantic Relations in a Categorical Verbal Fluency Test: An Exploratory Investigation in Mild Cognitive Impairment. *Frontiers in psychology*, *10*, 2797. https://doi.org/10.3389/fpsyg.2019.02797

Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K., & Kaye, J. (2011). Spoken Language Derived Measures for Detecting Mild Cognitive Impairment. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, 2081-2089. https://ieeexplore.ieee.org/document/5710404

Salthouse, T. (1996). The processing speed theory of adult age differences in cognition. Psychological Review, 103, 403-428. https://doi.apa.org/doi/10.1037/0033-295X.103.3.403

Szatloczki, G., Hoffmann, I., Vincze, V., Kalman, J., & Pakaski, M. (2015). 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. https://doi.org/10.3389/fnagi.2015.00195

Tovar, A., Garí Soler, A., Ruiz-Idiago, J., Mareca Viladrich, C., Pomarol-Clotet, E., Rosselló, J., & Hinzen, W. (2020). Language disintegration in spontaneous speech in Huntington's disease: a more fine-grained analysis. *Journal of communication disorders*, *83*, 105970. https://doi.org/10.1016/j.jcomdis.2019.105970

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037//0894-4105.11.1.138

Wang, Y., & Gennari, S. P. (2019). How language and event recall can shape memory for time. *Cognitive psychology*, *108*, 1–21. https://doi.org/10.1016/j.cogpsych.2018.10.003

Warkentin, S., Erikson, C., & Janciauskiene, S. (2008). rCBF pathology in Alzheimer's disease is associated with slow processing speed. Neuropsychologia, 46, 1193-1200. https://psycnet.apa.org/doi/10.1016/j.neuropsychologia.2007.08.029

Warren, D. E., Rubin, R., Shune, S., & Duff, M. C. (2018). Memory and language in aging: How their shared cognitive processes, neural correlates, and supporting mechanisms change with age. In M. Rizzo, S. Anderson, & B. Fritzsch (Eds.), The Wiley handbook on the aging mind and brain (pp. 270–295). Wiley Blackwell. https://doi.org/10.1002/9781118772034.ch14

Win, K. T., Pluta, J., Yushkevich, P., Irwin, D. J., McMillan, C. T., Rascovsky, K., Wolk, D., & Grossman, M. (2017). Neural Correlates of Verbal Episodic Memory and Lexical Retrieval in Logopenic Variant Primary Progressive Aphasia. *Frontiers in neuroscience*, *11*, 330. https://doi.org/10.3389/fnins.2017.00330

Zellner, B. (1994). Pauses and the Temporal Structure of Speech. *Fundamentals of Speech Synthesis and Speech Recognition*, 41–62.

Zimmerer, V. C., Varley, R. A., Deamer, F., & Hinzen, W. (2019). Factive and counterfactive interpretation of embedded clauses in aphasia and its relationship with lexical, syntactic and general cognitive capacities. *Journal of Neurolinguistics, 49,* 29–44. https://doi.org/10.1016/j.jneuroling.2018.08.002