The Role of Silence in Verbal Fluency Tasks – A New Approach for the Detection of Mild Cognitive Impairment

Réka Balogh[1,*] [ORCID] , Nóra Imre[1], Gábor Gosztolya[2], lldikó Hoffmann[3,4], Magdolna Pákáski[1] and János Kálmán[1]

[1]Department of Psychiatry, University of Szeged, Szeged, Hungary
[2]ELRN-SZTE Research Group on Artificial Intelligence, Szeged, Hungary
[3]Department of Hungarian Linguistics, University of Szeged, Szeged, Hungary
[4]Hungarian Research Centre for Linguistics, ELRN, Budapest, Hungary

Abstract

Objective: Most recordings of verbal fluency tasks include substantial amounts of task-irrelevant content that could provide clinically valuable information for the detection of mild cognitive impairment (MCI). We developed a method for the analysis of verbal fluency, focusing not on the task-relevant words but on the silent segments, the hesitations, and the irrelevant utterances found in the voice recordings. **Methods:** Phonemic ('k', 't', 'a') and semantic (animals, food items, actions) verbal fluency data were collected from healthy control (HC; $n = 25$; $M_{age} = 67.32$) and MCI ($n = 25$; $M_{age} = 71.72$) participants. After manual annotation of the voice samples, 10 temporal parameters were computed based on the silent and the task-irrelevant segments. Traditional fluency measures, based on word count (correct words, errors, repetitions) were also employed in order to compare the outcome of the two methods. **Results:** Two silence-based parameters (the number of silent pauses and the average length of silent pauses) and the average word transition time differed significantly between the two groups in the case of all three semantic fluency tasks. Subsequent receiver operating characteristic (ROC) analysis showed that these three temporal parameters had classification abilities similar to the traditional measure of counting correct words. **Conclusion:** In our approach for verbal fluency analysis, silence-related parameters displayed classification ability similar to the most widely used traditional fluency measure. Based on these results, an automated tool using voiced-unvoiced segmentation may be developed enabling swift and cost-effective verbal fluency-based MCI screening.

Keywords: Cognitive aging, Mild cognitive impairment, Neuropsychology, Verbal fluency, Semantic memory, Speech parameters

INTRODUCTION

Mild cognitive impairment (MCI) is a heterogeneous clinical syndrome, often considered a transitional stage between healthy cognitive aging and dementia (Petersen, 2004), and it is also associated with an increased risk of developing dementia later on (Roberts et al., 2014). Early recognition and timely diagnosis are crucial in MCI, because they can provide an opportunity to reduce the rate of cognitive decline (Hahn & Andel, 2011), while also offering a chance for the patients and their relatives to start planning for the future (Knopman & Petersen, 2014). Considering the high prevalence of MCI (Roberts & Knopman, 2013) and especially the constantly overburdened clinical settings, it would be beneficial to replace the current labor-intensive and time-consuming assessments of cognitive functioning with swift, low-cost, and preferably automated tools.

Verbal fluency tests are neuropsychological tests, extensively used both in research and in the clinical practice. In the standard versions of the fluency tests, participants are given 60 s to list as many words as they can, beginning with a given letter (phonemic fluency) (Borkowski, Benton, & Spreen, 1967) or belonging to a given semantic category (semantic fluency) (Newcomb, 1969). There is an additional, third type of verbal fluency task: action fluency (or verb fluency), where the patients have to produce as many verbs ('things that people do') as they can (Piatt, Fields, Paolo, & Troster, 1999). However, in the current study, for the sake of simplicity, action fluency will be regarded as a semantic fluency task, because both semantic fluency and action fluency are content-oriented speech tasks (Östberg, Fernaeus, Hellstrom, Bogdanovic, & Wahlund, 2005).

*Correspondence and reprint requests to: Réka Balogh, Department of Psychiatry, University of Szeged, Korányi Fasor 8-10, Szeged, H-6720, Hungary. E-mail: balogh.reka@med.u-szeged.hu

Both phonemic and semantic fluency tasks require rapid associative exploration; however, semantic fluency relies more on semantic associations and reflects more on the integrity of semantic memory. On the other hand, phonemic fluency depends more on search strategies based on lexical representation (Henry, Crawford, & Phillips, 2004; Teng et al., 2013). Executive control processes also play a major role in the execution of verbal fluency tests, because during the task, subjects not only need to remember the exact instruction and keep the already used responses in mind, but they must also repress the repetitions and other potentially incorrect or irrelevant responses (Shao, Janse, Visser, & Meyer, 2014). Fluency tests have been validated in the assessment of verbal and executive skills (Shao et al., 2014), and both of these abilities have been reported to deteriorate in dementia and in other forms of cognitive impairments. Therefore, fluency tests have a great potential as effective screening tools for MCI (García-Herranz, Diaz-Mardomingo, Venero, & Peraita, 2020; McDonnell et al., 2020).

The traditional, most common approach for the assessment of verbal fluency performance requires the clinician to count the number of unique and correct words, along with the number of errors and the number of repetitions produced by the participant. This analysis can be refined by scoring the number of correct words based on time intervals (e.g., 0–20, 21–40, 41–60 s) (Demetriou & Holtzer, 2017; Jacobs, Mercuri, & Holtzer, 2021). Moving beyond simple word counts, a more sophisticated, qualitative method can be applied, which is called clustering. In this method, consecutive words are clustered based on linguistic similarity or a shared category (e.g., rhyming words in the case of phonemic fluency tasks, or pets in the case of the animal fluency task). Thus, the average sizes of the clusters and the number of switches between these clusters can be examined (Troyer, Moscovitch, & Winocur, 1997). Even though this approach may provide more information about the underlying mental processes, it is also relatively time-consuming. Furthermore, compared to the most widespread, word count-based assessment, this method requires the manual coding and grouping of words, which may even raise reliability issues (Taler, Johns, & Johns, 2020).

Recently, there have been multiple attempts with different approaches to overcome the disadvantages of the above-mentioned methods by introducing automated analyses. These approaches have the benefit of being objective, repeatable, and they also yield quick output (König et al., 2018). The majority of these methods focus on the computation and analysis of semantic clusters. Latent semantic analysis (LSA) can be applied to examine the strength of the semantic relationship of two consecutive words by constructing a co-occurrence matrix for all of the words found in a given corpus of text (Ledoux et al., 2014; Pakhomov & Hemmy, 2014). A more recent computational method, called explicit semantic analysis (ESA), examines Wikipedia entries for the quantification of relationships between words based on different types of similarities (e.g., taxonomic, geographic, or linguistic) (Woods, Wyma, Herron, & Yund, 2016). It is also possible to combine semantic measures with temporal information. In this approach, the recalled words are organized in clusters defined semantically and also in clusters based on the temporal proximity of the words (Tröger et al., 2019). Verbal fluency tasks can also be analyzed by exploring certain speech features that can be automatically extracted from fluency voice recordings (Lopez-de-Ipina et al., 2015).

However, there is a major obstacle in the application of the automatic analysis of fluency recordings that stems from the general characteristics of the responses produced by the participants: most voice recordings of fluency test performances contain more than just a sequence of task-relevant words. The recordings also contain speech segments irrelevant in terms of the task, including filler words or hesitations, irrelevant comments, questions directed at the examiner, or loud thinking. To be able to automatically analyze the task-relevant words, fluency recordings need to go through a time-consuming preparation process prior to the analysis: the words irrelevant to the task need to be removed from the recording or transcript, and some words need to be lemmatized (i.e., converted to their stem) (Chen et al., 2020; Holmlund, Cheng, Foltz, Cohen, & Elvevag, 2019).

Given the substantial amount of task-irrelevant content in most fluency recordings, the question arises whether the analysis of these segments could provide valuable information regarding the overall verbal fluency performance of the patient. After manually annotating the recordings, we derived temporal parameters that, instead of targeting the task-relevant words, contained the silent segments, the hesitations, and the utterances irrelevant to the task. Therefore, the focus of this exploratory study was to move beyond the words recalled by the participants and explore the additional, previously unharvested information present in the fluency recordings. It should be noted that this approach, similarly to the previously summarized methods, required substantial manual work. However, in the future (depending on the characteristics of the given parameter) it could allow the development of automatic analysis.

Our main goal was: (1) to examine whether these parameters can differentiate between participants classified as healthy control (HC) and as MCI (*temporal analysis method*). Besides the temporal parameters, traditional fluency scores (number of correct words, errors, and repetitions) were also calculated for the same fluency recordings (*traditional analysis method*). We sought; (2) to compare the two methods of analysis regarding their ability to detect differences in the performance of the HC and MCI groups. The inclusion of both phonemic and semantic fluency tasks in the research protocol also allowed us; and (3) to compare the different types of fluency tasks to investigate their sensitivity to the presence of MCI.

METHODS

Participants

Participants (patients and their relatives, scheduled for consultations) were recruited at the Memory Clinic of the Department of Psychiatry, University of Szeged (Szeged,

Hungary). Data collection was carried out between February 2018 and March 2020.

The required sample size for the study was assessed a priori using G * Power v.3.2.9.7. (Faul, Erdfelder, Lang, & Buchner, 2007) with the settings of effect size $d = 0.8$; alpha error probability: 0.05, power (1-beta error probability): 0.8. Based on this, the optimal sample size was calculated as 52, which later (due to COVID-19 regulations halting data collection in clinical research) was limited to 50. Initially, a total of 79 individuals were recruited to take part in the study.

Inclusion criteria included at least 50 years of age, a minimum of 8 years of formal education, and Hungarian as a native language. Individuals were excluded if they had any past or present neuropsychological, psychotic or mood disorders, head injuries, stroke, substance abuse disorders, major (uncorrected) hearing loss, or language problems (e.g., stutter), based on patient history and medical records. Participants with MRI or CT records showing evidence of micro- or macrohemorrhages, lacunar or other infarctions, cerebral contusion, encephalomalacia, aneurysm, vascular malformation, or space-occupying lesions were also excluded.

In addition, the two main exclusion criteria were the presence of dementia or major cognitive deficits and depression. To rule out possible cases of dementia, the Mini-Mental State Examination (MMSE) (Folstein, Folstein, & Mchugh, 1975) was applied as a screening tool: participants with a score of 24 or below were excluded from the study. The possibility of depression was assessed using the 15-item version of the Geriatric Depression Scale (GDS-15) (Yesavage & Sheikh, 1986): participants scoring 7 or above on the test were excluded. After reviewing and evaluating the criteria, 50 subjects were considered eligible for inclusion in the study (Figure 1).

Participants were split into two groups based on their MMSE scores. MMSE cut-off scores were determined based on the results of previous studies conducted by our research group: in these works, the mean scores of MMSE emerged as $29.17 \pm 0.71/29.24 \pm 0.523$ for the HC and $26.97 \pm 0.96)/27.16 \pm 0.898$ for the MCI group (Gosztolya et al., 2019; Toth et al., 2018). Hence, participants achieving a score of 29 to 30 points were considered as healthy control (HC) subjects, while participants achieving a score of 25 to 28 points formed the MCI group. The subtypes of MCI (amnestic or non-amnestic) were not considered. The two groups showed no significant difference in gender and years of education. However, participants of the MCI group were significantly older than the participants enrolled in the HC group. No significant difference was found in the GDS-15 score between the two groups (Table 1).

Study Protocol

Each participant performed a series of neuropsychological tests: six fluency tasks, the Digit Span Test – Forward and Backward (Wechsler, 1981), the Non-Word Repetition Test (Gathercole, Willis, Baddeley, & Emslie, 1994), the Listening Span Test (Daneman & Carpenter, 1980), the Clock Drawing Test (Shulman, Shedletsky, & Silver,

1986) and the Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog) (Rosen, Mohs, & Davis, 1984). The fluency tasks were implemented in a fixed order, separated by the five shorter cognitive tests, while ADAS-Cog was administered at the very end of the study protocol to prevent fatigue. We also ensured that tasks assessing the same cognitive domain did not follow each other directly.

In the three phonemic fluency tasks, the participants were asked to list as many words as they can, starting with the letters 'k', 't', and 'a', respectively, while avoiding proper nouns. For the semantic fluency tasks, participants had to name as many animals, food items, and actions (verbs – 'things that people do') as they could. The participants were instructed to avoid saying variations of the same word stem (e.g., horse, horses; go, goes). For all 6 verbal fluency tasks, participants had 1 min to perform the task. The 1-min interval began with the investigator saying: 'Start.' Every verbal fluency task was recorded using an Olympus Digital Voice Recorder (16 kHz sampling rate, 16-bit resolution). The recordings were also transcribed manually for the calculation of the traditional scores. Therefore, fluency performances were analyzed in two ways: by implementing the novel temporal parameters, and also by using the traditional method, based on word count.

Analysis Method Based on Temporal Parameters

Manual transcription process of the fluency recordings

Voice recordings of all fluency tasks were manually transcribed in Praat, a free language software enabling speech analysis (Boersma & Weenink, 2020). The transcription process was supervised by a linguist specialized in language pathologies (I. H.), while quality control was ensured by an expert in the field of computational speech processing (G. G.). Due to the quality of their recordings, an HC participant's animal category fluency task and an MCI participant's 'k' letter fluency task were unsuitable for transcription; therefore, these recordings were not considered in the analysis of temporal parameters, but they were included in the traditional analysis.

Annotation of speech features in the verbal fluency recordings

The transcriptions of the fluency recordings contained not only the task-relevant answers of the participants (the recalled words – including correct, incorrect, and repeated words), but also silent pauses, hesitation sounds (filled pauses, like 'hmm' and 'er'), and irrelevant utterances, such as comments or loud thinking said by the subjects (e.g., 'did I say this before?', 'uh, it's not an easy task, let me think . . .'). False starts ('te-. . . tiger'), as well as laughing and coughing sounds were also annotated. The laughing, coughing, and false starts parameters were considered unintentional and
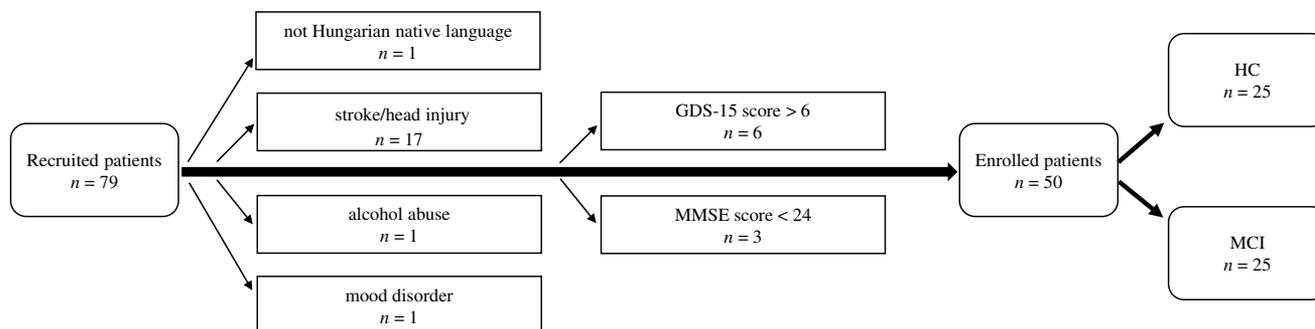
Fig. 1. Flowchart of the participant exclusion process. (GDS-15:15-item Geriatric Depression Scale; MMSE: Mini-Mental State Examination; HC: healthy control; MCI: mild cognitive impairment).

Table 1. Descriptive and comparative statistics for the demographic characteristics and neuropsychological test scores of the study participants

	HC ($n = 25$)	MCI ($n = 25$)		
	M (SD)		Comparative test statistics	*p*
Demographics				
Gender (male/female)	8/17	7/18	$\chi^2(1) = 0.095$	0.758
Age (years)	67.32 (8.300)	71.72 (5.435)	$U = 187.000; Z = -2.440$	**0.015**
Education (years)	13.48 (2.632)	12.36 (2.827)	$U = 255.500; Z = -1.136$	0.256
Neuropsychological test scores				
MMSE	29.44 (0.507)	26.96 (1.060)	$U = 0.000; Z = -6.202$	**< 0.001**
GDS-15	1.84 (1.724)	2.40 (1.225)	$U = 232.500; Z = -1.587$	0.112

HC: healthy control; MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; GDS-15: 15-item Geriatric Depression Scale.
Significant *p*-values ($p < 0.05$) are in **bold.**

were discarded from further analysis because we found that the number of these occurrences was negligible.

Calculation of temporal parameters based on the speech features

For each recording, task-relevant words, silent segments, hesitation sounds, and irrelevant utterances were annotated based on their boundaries (their exact start and end times), providing their duration measures. Based on this, *the total number, the average length,* and *the total length of silent pauses*; *the total number, the average length,* and *the total length of hesitations*; and *the total number, the average length, and the total length of irrelevant utterances* were calculated. Besides these parameters, the mean time between two consecutive task-relevant words (*average word transition time*) was also calculated based on the transcript. Not only correct words but also the errors and repetitions were considered task-relevant words. The average word transition time, irrelevant of its content, such as silent pause, hesitation, or irrelevant utterance, provided information about the average time the participant needed to produce a new task-relevant word, and because of this, it had a positive association with the average and total length of silent pauses, hesitations, and irrelevant utterances.

It is worth noting that because of the distinctive regular rhythm that is inherent in verbal fluency performances, each

of the task-relevant words listed by the participants was separated by a silent pause (irrelevant of its length). Consequently, the number of silent pauses increased in parallel with the number of task-relevant words said by the participant. Therefore, analyzing the number of silent pauses can be viewed as the converse of the traditional approach of counting only the task-relevant words.
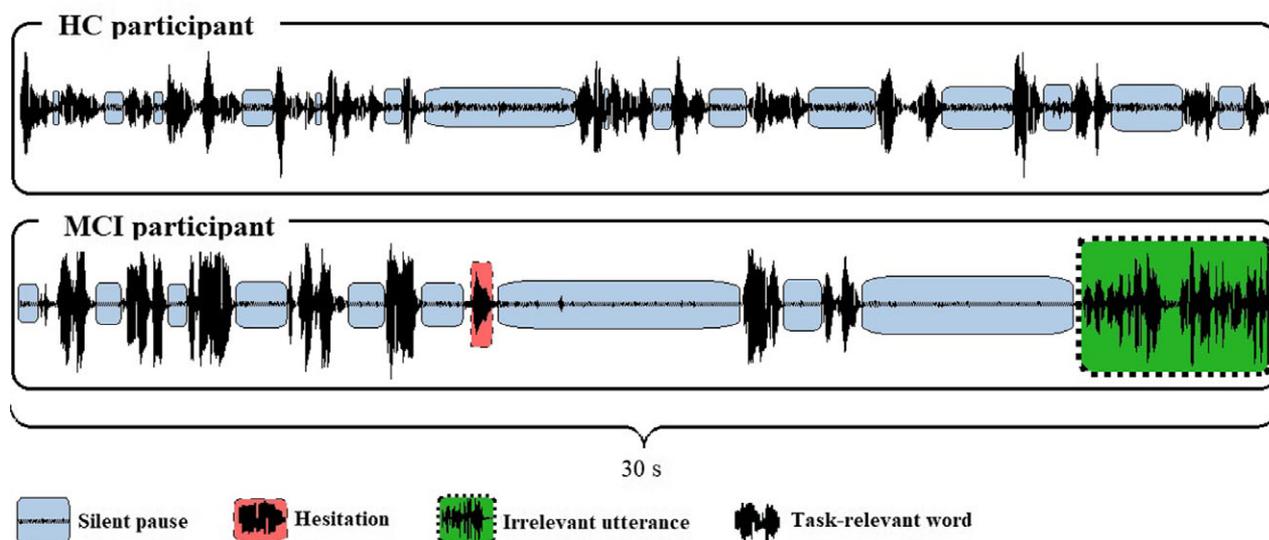
The parameters used in the study are listed and defined in Table 2; two waveform extracts from a fluency task performed by an HC and an MCI subject are shown in Figure 2.

Traditional Fluency Analysis Based on Word Count

In the traditional scoring method (Lezak, 2012), we calculated the number of correct words, the number of errors, and the number of repetitions or perseverations; the last two were considered as one variable. In the case of animal fluency, when a participant recalled synonymous words (e.g., cat and kitten), variations in gender (e.g., hen and rooster), or an animal and its offspring (e.g., horse and foal), words were only scored as one. The participants did not receive points for naming a subcategory if they also gave specific examples of it [e.g., in the case of food items: fruit (0 points), apple (1 point), pear (1 point)].

Table 2. List and definitions of the temporal parameters

Temporal fluency parameters	Description
Silent pause parameters	
Total number of silent pauses (count)	Number of silent segments
Average length of silent pauses (s)	Average length of silent segments
Total length of silent pauses (s)	Total length of silent segments
Hesitation parameters	
Total number of hesitations (count)	Total number of filled pauses (e.g., 'hmm', 'umm')
Average length of hesitations (s)	Average length of filled pauses (e.g., 'hmm', 'umm')
Total length of hesitations (s)	Total length of filled pauses (e.g., 'hmm', 'umm')
Irrelevant utterances parameters	
Total number of irrelevant utterances (count)	Total number of filler words and comment blocks (including articles and conjunctions)
Average length of irrelevant utterances (s)	Average length of filler words and comment blocks (including articles and conjunctions)
Total length of irrelevant utterances (s)	Total length of filler words and comment blocks (including articles and conjunctions)
Average word transition time (s)	Mean period of time between two consecutive 'task-oriented' words

Fig. 2. Waveforms extracted from the food item fluency recordings of two participants. (Extracted from Praat. HC: healthy control; MCI: mild cognitive impairment).

Statistical Analysis

Descriptive statistical analysis was used to examine the demographic features, the neuropsychological test scores, and the fluency measures of the participants. The assumption of normality was not met according to the results of the Shapiro–Wilk test in more than two-thirds of the cases, therefore, in order to obtain comparable statistical measures, comparisons between the HC and the MCI groups were executed using the Mann–Whitney U test. Categorical variables were compared using the Chi-square test. Effect sizes were calculated using the Pearson correlation coefficient ($r = \frac{z}{\sqrt{N}}$) (Rosenthal, 1991).

Receiver operating characteristic (ROC) analysis was applied to assess the classification abilities of the temporal parameters and the traditional scores. Sensitivity and specificity were calculated using threshold values that yielded the highest possible sensitivity (while keeping specificity at

a minimum of 50%). For the comparison of classification abilities, the differences between the area under the curve variables (AUCs) were compared based on the method of DeLong, DeLong, and Clarke-Pearson (1988).

For all statistical comparisons, the level of significance was set at $p < 0.05$. All analyses were performed using SPSS v.24 (IBM SPSS Statistics for Windows, 2016), except for the comparison of AUCs, for which the MedCalc Statistical Software v.19.6. (MedCalc Software, 2020) was utilized.

RESULTS

Temporal Parameters of Verbal Fluency Performance

Considering the phonemic fluency tasks, in the 'a' fluency, the average length and the total length of irrelevant utterances

were significantly higher in the MCI group, while none of the temporal parameters differed between the two groups in the case of the 'k' and 't' phonemic fluencies (Table 3). Regarding the three semantic fluencies, the total number of silent pauses were significantly higher in the HC group in the animal and action fluency tasks, whereas the average length of silent pauses and the average word transition time were significantly higher in the MCI group throughout all of the three tasks (Table 4).

Traditional Word Count Measures of Verbal Fluency Performance

In the three phonemic fluency tasks, no statistically significant difference was found between the groups regarding the number of correct words and the number of repetitions or perseverations. However, in the 'a' phonemic fluency task, participants from the MCI group produced more errors than participants from the HC group (Table 5). As for the semantic fluency tests, participants from the HC group had a significantly higher number of correct words in the case of all three (animals, food items, and actions) tasks. In the number of repetitions or perseverations, there was no statistically significant difference between the two study groups (Table 6).

ROC Analysis of the Significant Temporal Parameters

ROC analysis of the temporal parameters was carried out in the case of the five parameters that, based on the previously conducted comparative tests, showed significant differences between the HC and MCI groups.

The analysis revealed that the average length and the total length of irrelevant utterances had a significant classification ability in the case of the 'a' phonemic fluency, with the same sensitivity (80%) and specificity (52%) for both parameters. In the semantic fluency tests, the number of silent pauses had significant classification ability in both the animal and action fluency tests, while the average length of silent pauses and the average word transition time was shown to be able to discriminate between the groups in the case of all three semantic fluency tests. Sensitivity was the highest in the case of the average word transition time in the animal fluency test (sensitivity: 96.0%; specificity: 62.5%). Accuracy measures of the temporal parameters that differed between the groups are given in Table 7. For every ROC analysis, sensitivity and specificity were determined using threshold values optimal for early screening, i.e., maximizing the sensitivity, while keeping specificity greater than or equal to 50%.

ROC Analysis of the Significant Traditional Measures

ROC analysis was also executed on the traditional measures that showed significant differences between the HC and MCI groups, to determine the classification ability of these measures. The analysis revealed that the number of errors in the 'a' phonemic fluency test had no significant classification ability. With respect to semantic fluency tests, the number of correct words showed significant classification abilities in the case of the animal, the food item, and the action fluencies. The animal naming fluency showed the highest sensitivity of 100% (specificity: 56%). Accuracy measures of the traditional fluency scores that showed significant differences between the groups are given in Table 8.

Comparison of the Temporal and Traditional Measures Regarding their Classification Ability

Pairwise comparisons of AUCs were executed to compare the classification ability of the three significant temporal parameters (total number of silent pauses, average length of silent pauses, average word transition time) and the significant traditional measure (number of correct words) in the semantic fluency tasks. In the animal category fluency, the results indicated no significant differences regarding AUCs between the number of correct words and the total number of silent pauses ($Z = 1.433$, $p = 0.151$) or the average word transition time ($Z = 1.579$, $p = 0.114$), however, the classification ability of the average length of silent pauses was smaller ($Z = 2.043$, $p = 0.041$) compared to the correct word count. In the case of the food item fluency, no difference was found between the AUCs of the number of correct words and the average length of silent pauses ($Z = 0.978$, $p = 0.328$), and the average word transition time ($Z = 0.662$, $p = 0.508$). Furthermore, in action fluency, the classification ability of correct word count did not differ from either the total number of silent pauses ($Z = 0.267$, $p = 0.789$), the average length of silent pauses ($Z = 0.056$, $p = 0.954$) or the average word transition time ($Z = 0.046$, $p = 0.962$).

DISCUSSION

Main Findings

This study presents a new practical framework for verbal fluency analysis. To the best of our knowledge, we are the first to report on verbal fluency performance beyond the recalled words, focusing on the pauses and task-irrelevant content of speech in the fluency recordings. We quantitatively analyzed a number of temporal parameters that were calculated based on silent pauses, hesitations, and irrelevant speech segments annotated in the recordings. Our main finding is that in the case of semantic fluency tests, some of the temporal parameters based on silent pauses can discriminate between individuals with cognitive impairment and individuals with healthy cognition. These results suggest that the analysis of these temporal parameters may complement or even substitute the widely applied, but more time-consuming and labor-intensive traditional word scoring method, while still providing comparable classification ability.

Table 3. Descriptive measures and statistical comparison of the temporal parameters in the phonemic fluency tasks

Phonemic fluency tasks	HC	MCI	Mann–Whitey U test			Effect size[r]
Temporal parameters	M (SD)		U	Z	p	r
Letter 'k'	$n = 25$	$n = 24^*$				
Total number of silent pauses (count)	19.040 (4.485)	17.291 (4.591)	230.500	-1.394	0.163	0.19
Average length of silent pauses (s)	2.438 (0.941)	2.767 (1.031)	244.000	-1.120	0.263	0.16
Total length of silent pauses (s)	42.569 (6.571)	43.532 (5.688)	278.000	-0.440	0.660	0.06
Total number of hesitations (count)	2.000 (2.645)	1.708 (2.095)	281.500	-0.382	0.702	0.05
Average length of hesitations (s)	0.482 (0.448)	0.398 (0.382)	279.500	-0.421	0.674	0.06
Total length of hesitations (s)	1.350 (1.737)	1.137 (1.334)	283.000	-0.349	0.727	0.05
Total number of irrelevant utterances (count)	3.280 (4.559)	4.333 (3.818)	225.000	-1.517	0.129	0.22
Average length of irrelevant utterances (s)	1.021 (0.666)	1.242 (0.851)	274.000	-0.520	0.603	0.07
Total length of irrelevant utterances (s)	4.283 (7.149)	4.889 (3.609)	213.000	-1.742	0.082	0.25
Average word transition time (s)	4.505 (2.687)	5.159 (3.979)	230.000	-1.400	0.162	0.20
Letter 't'	$n = 25$	$n = 25$				
Total number of silent pauses (count)	18.320 (4.269)	16.920 (5.259)	257.000	-1.081	0.280	0.15
Average length of silent pauses (s)	2.521 (0.879)	2.993 (1.542)	261.000	-0.999	0.318	0.14
Total length of silent pauses (s)	42.847 (5.770)	43.834 (5.666)	278.000	-0.669	0.503	0.07
Total number of hesitations (count)	1.480 (2.023)	1.720 (2.051)	290.000	-0.455	0.649	0.06
Average length of hesitations (s)	0.520 (0.509)	0.443 (0.348)	293.500	-0.375	0.708	0.05
Total length of hesitations (s)	1.128 (1.504)	1.069 (1.326)	312.500	0.000	1.000	0.00
Total number of irrelevant utterances (count)	3.240 (3.562)	3.720 (2.806)	256.500	-1.097	0.273	0.16
Average length of irrelevant utterances (s)	0.967 (0.580)	1.228 (0.616)	231.500	-1.573	0.116	0.21
Total length of irrelevant utterances (s)	4.154 (5.656)	4.825 (3.379)	234.500	-1.515	0.130	0.21
Average word transition time (s)	3.816 (1.739)	4.944 (3.045)	250.000	-1.213	0.225	0.17
Letter 'a'	$n = 25$	$n = 25$				
Total number of silent pauses (count)	13.920 (3.639)	14.120 (4.876)	298.000	-0.283	0.778	0.04
Average length of silent pauses (s)	3.636 (1.446)	3.853 (2.834)	268.000	-0.863	0.388	0.12
Total length of silent pauses (s)	45.881 (5.219)	43.042 (7.551)	235.00	-1.504	0.133	0.01
Total number of hesitations (count)	1.040 (1.059)	1.200 (1.354)	311.000	-0.031	0.976	0.00
Average length of hesitations (s)	0.640 (0.565)	0.462 (0.485)	263.500	-0.974	0.330	0.14
Total length of hesitations (s)	0.973 (1.316)	0.985 (1.254)	301.500	-0.219	0.827	0.03
Total number of irrelevant utterances (count)	3.480 (4.154)	4.560 (3.292)	214.500	-1.918	0.055	0.27
Average length of irrelevant utterances (s)	1.065 (0.701)	1.630 (0.725)	180.000	-2.572	**0.010**	0.36
Total length of irrelevant utterances (s)	4.637 (5.286)	7.160 (5.322)	204.000	-2.106	**0.035**	0.30
Average word transition time (s)	5.115 (2.651)	5.224 (2.839)	286.000	-0.514	0.607	0.07

M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment.
Significant p-values ($p < 0.05$) are in **bold**.
*One fluency voice recording was unsuitable for transcription.
[r] Effect size is calculated as Pearson's r, expressed in absolute value.
Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

Three temporal parameters (total number of silent pauses, average length of silent pauses, and average word transition time) consistently differed between the HC and MCI groups in the case of the semantic (animal, food item, and action) fluency tests. In the phonemic fluency tests, differences could only be observed in the case of the 'a' phonemic fluency, where the average and total lengths of irrelevant utterances showed significant differences.

It should be noted that the direction of differences in the silence-based parameters might seem inconsistent: the average lengths of the silent pauses and the average word transition times were longer in the MCI group, whereas HC participants had a higher number of silent pauses in the case of the semantic tasks. Since silent pauses were defined as the absence of speech/sound regardless of length, every detectable silent segment found in the recordings was annotated as a silent pause, including even the brief transitions between words. Therefore, the number of silent pauses was increased by the number of words uttered by the participant. Since the HC group produced significantly more correct words in semantic fluency tasks, the number of silent pauses was also significantly higher in this group.

Table 4. Descriptive measures and statistical comparison of the temporal parameters in the semantic fluency tasks

Semantic fluency tasks	HC	MCI	Mann–Whitney *U* test			Effect size[r]
Temporal parameters	*M* (*SD*)		*U*	*Z*	*p*	*r*
Animals	*n* = 24*	*n* = 25				
Total number of silent pauses (count)	25.666 (4.603)	21.760 (4.968)	156.000	-2.890	**0.004**	0.41
Average length of silent pauses (s)	1.437 (0.445)	1.883 (0.718)	179.000	-2.420	**0.016**	0.34
Total length of silent pauses (s)	35.489 (6.485)	37.982 (8.193)	229.000	-1.420	0.156	0.20
Total number of hesitations (count)	3.166 (2.371)	3.240 (3.620)	271.000	-0.586	0.558	0.08
Average length of hesitations (s)	0.564 (0.290)	0.460 (0.358)	237.500	1.255	0.209	0.18
Total length of hesitations (s)	2.195 (1.982)	2.139 (2.820)	264.500	-0.713	0.476	0.10
Total number of irrelevant utterances (count)	3.333 (3.595)	5.120 (4.850)	231.500	-1.380	0.167	0.20
Average length of irrelevant utterances (s)	1.019 (0.641)	1.146 (0.727)	277.000	-0.461	0.645	0.07
Total length of irrelevant utterances (s)	4.379 (6.116)	6.562 (5.647)	220.000	-1.603	0.109	0.23
Average word transition time (s)	2.021 (0.756)	2.852 (0.841)	128.000	-3.440	**0.001**	0.49
Food items	*n* = 25	*n* = 25				
Total number of silent pauses (count)	25.400 (6.062)	21.720 (5.926)	216.000	-1.877	0.061	0.26
Average length of silent pauses (s)	1.395 (0.504)	1.888 (0.937)	201.000	-2.163	**0.031**	0.30
Total length of silent pauses (s)	33.192 (6.464)	36.368 (7.200)	242.000	-1.368	0.171	0.19
Total number of hesitations (count)	2.600 (2.432)	2.600 (2.661)	307.000	-0.109	0.913	0.02
Average length of hesitations (s)	0.444 (0.348)	0.494 (0.435)	306.000	-0.128	0.898	0.02
Total length of hesitations (s)	1.636 (1.544)	1.855 (2.015)	302.000	-0.207	0.836	0.03
Total number of irrelevant utterances (count)	3.600 (3.905)	4.360 (4.733)	294.000	-0.362	0.717	0.05
Average length of irrelevant utterances (s)	0.772 (0.581)	1.051 (1.028)	273.000	-0.770	0.441	0.11
Total length of irrelevant utterances (s)	3.716 (4.898)	5.210 (5.353)	259.000	-1.044	0.297	0.15
Average word transition time (s)	1.755 (0.770)	2.630 (1.356)	171.000	-2.746	**0.006**	0.40
Actions	*n* = 25	*n* = 25				
Total number of silent pauses (count)	24.240 (6.332)	19.080 (5.597)	184.000	-2.502	**0.012**	0.35
Average length of silent pauses (s)	1.600 (0.565)	2.373 (1.439)	192.000	-2.338	**0.019**	0.33
Total length of silent pauses (s)	35.898 (5.605)	38.524 (7.485)	230.000	-1.601	0.109	0.22
Total number of hesitations (count)	2.720 (2.282)	2.840 (2.511)	309.000	-0.069	0.945	0.01
Average length of hesitations (s)	0.547 (0.362)	0.554 (0.477)	292.000	-0.401	0.689	0.06
Total length of hesitations (s)	1.963 (1.741)	2.096 (2.290)	302.000	-0.205	0.837	0.03
Total number of irrelevant utterances (count)	4.040 (3.920)	4.160 (3.681)	307.500	-0.098	0.922	0.01
Average length of irrelevant utterances (s)	1.069 (0.626)	1.153 (0.760)	290.500	-0.427	0.669	0.06
Total length of irrelevant utterances (s)	4.302 (4.600)	5.188 (4.351)	273.500	-0.757	0.449	0.11
Average word transition time (s)	2.258 (0.996)	2.989 (1.199)	196.000	-2.260	**0.024**	0.32

M: mean; *SD*: standard deviation; HC: healthy control; MCI: mild cognitive impairment.
Significant *p*-values (*p* < 0.05) are in **bold**.
*One fluency voice recording was unsuitable for transcription.
[r] Effect size is calculated as Pearson's *r*, expressed in absolute value.
Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

The average word transition time parameter also had a direct influence on the number of correct words. Since this parameter contains every task-irrelevant segment, the increase of the average word transition time by definition led to the decrease of the number of recalled words, therefore it could be viewed that these two parameters were somewhat inversely proportional. The average length of silent pauses parameter also affected the number of correctly recalled words. However, this is less of a general phenomenon, since the average length of silent pauses does not have a sole effect on the number of recalled words – it can be also significantly influenced by other task-irrelevant contents of speech (e.g., loud hesitations).

The importance of silent pauses has also been highlighted in the area of connected speech analysis: studies have shown that compared to HC subjects, participants with MCI produce more and longer silent pauses in their speech (Sluis et al., 2020; Toth et al., 2018). Even though spontaneous speech samples provide ecologically valid data, utilizing verbal fluency tests for the analysis of speech can be even more advantageous, as it can be combined with already standardized qualitative approaches. To be able to compare the results

Table 5. Descriptive measures and statistical comparison of the traditional fluency scores in the phonemic fluency tests

Traditional fluency scores of the phonemic fluency tasks	HC M (SD) n = 25	MCI M (SD) n = 25	Mann–Whitney U test U	Z	p	Effect size[r] r
Letter 'k'						
Correct words	13.68 (4.571)	11.52 (4.700)	227.000	-1.667	0.096	0.24
Errors	0.04 (0.200)	0.16 (0.374)	275.000	-1.400	0.162	0.20
Repetitions/perseverations	0.16 (0.374)	0.32 (0.690)	294.000	-0.537	0.591	0.08
Letter 't'						
Correct words	12.88 (4.314)	10.76 (4.371)	233.000	-1.547	0.122	0.22
Errors	0.20 (0.408)	0.28 (0.614)	307.500	-0.139	0.889	0.02
Repetitions/perseverations	0.48 (0.653)	0.28 (0.678)	248.500	-1.577	0.115	0.22
Letter 'a'						
Correct words	8.68 (3.424)	7.32 (3.987)	240.000	-1.416	0.157	0.20
Errors	0.12 (0.332)	0.72 (1.208)	231.500	-2.106	**0.035**	0.30
Repetitions/perseverations	0.20 (0.577)	0.20 (0.408)	292.500	-0.609	0.542	0.09

M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment).
Significant p-values (p < 0.05) are in **bold**.
[r] Effect size calculated as Pearson's r, expressed in absolute value. Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

Table 6. Descriptive measures and statistical comparison of the traditional fluency scores in the semantic fluency tests

Traditional fluency scores of the semantic fluency tasks	HC M (SD) n = 25	MCI M (SD) n = 25	Mann–Whitney U test U	Z	p	Effect size[r] r
Animals						
Correct words	20.54 (4.412)	14.76 (3.358)	99.000	-4.154	**0.000**	0.59
Errors	0.00 (0.000)	0.04 (0.200)	300.000	-1.000	0.317	0.14
Repetitions/perseverations	0.42 (0.584)	0.48 (0.963)	298.000	-0.343	0.731	0.05
Food items						
Correct words	22.72 (6.073)	17.16 (5.249)	156.500	-3.034	**0.002**	0.43
Errors	0.04 (0.200)	0.04 (0.200)	312.500	0.000	1.000	0.00
Repetitions/perseverations	0.28 (0.458)	0.40 (0.764)	311.000	-0.038	0.970	0.01
Actions						
Correct words	18.72 (6.175)	14.40 (4.916)	194.500	-2.293	**0.022**	0.32
Errors	0.04 (0.200)	0.04 (0.200)	312.500	0.000	1.000	0.00
Repetitions/perseverations	0.40 (0.764)	0.48 (0.918)	308.500	-0.098	0.922	0.01

M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment.
Significant p-values (p < 0.05) are in **bold**.
[r] Effect size calculated as Pearson's r, expressed in absolute value.
Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

of these two types of study, it is important to note the difference between connected (spontaneous) speech and verbal fluency performances. Compared to connected speech, where pauses appear more randomly, in the fluency recordings silent pauses (with varying lengths) appear between every word, therefore producing a 'word-pause-word-pause'-like sequence. Because of these distinct characteristics, the number of silent pauses needs to be interpreted based on the methodology of the specific study.

Most recent approaches to verbal fluency analysis usually focus on the semantic content when evaluating fluency performance (Tröger et al., 2019; Woods et al., 2016). In contrast, this work focused on the examination of more easily quantifiable, objective variables; nevertheless, we were able to achieve classification abilities comparable to those reported in previous studies [AUC: 0.758 (König et al., 2018), AUC: 0.77 (Chen et al., 2020)]. The significant classification ability of the silent pause parameters in our study suggests that differentiation between HC and MCI patients' semantic verbal fluency performance may be possible by examining only the silent pauses in their speech. This can be achieved, for example, by dividing the voice recordings into voiced and unvoiced segments (Lopez-de-Ipina et al., 2015).

Table 7. Accuracy measures of those temporal parameters that significantly differed between the two groups based on the previous comparative statistic tests

Fluency tasks	Temporal parameters	Accuracy measures					
		p	AUC	95% CI-	95% CI+	Sensitivity (%)	Specificity (%)
Letter 'a'	Average length of irrelevant utterances (s)	**0.010**	0.712	0.569	0.855	80.0	52.0
	Total length of irrelevant utterances (s)	**0.035**	0.674	0.523	0.824	80.0	52.0
Animals	Total number of silent pauses (count)	**0.004**	0.740	0.598	0.882	76.0	50.0
	Average length of silent pauses (s)	**0.016**	0.702	0.549	0–855	72.0	50.0
	Average word transition time (s)	**0.001**	0.787	0.651	0.922	96.0	62.5
Food items	Average length of silent pauses (s)	**0.031**	0.678	0.528	0.828	68.0	52.0
	Average word transition time (s)	**0.006**	0.726	0.587	0.866	76.0	52.0
Actions	Total number of silent pauses (count)	**0.013**	0.706	0.562	0.849	72.0	52.0
	Average length of silent pauses (s)	**0.019**	0.693	0.544	0.841	72.0	52.0
	Average word transition time (s)	**0.024**	0.686	0.536	0.837	80.0	52.0

AUC: area under the curve; CI: confidence interval.
Significant *p*-values (**$p < 0.05$**) indicate that the measure is significantly better than chance at discriminating individuals of the two groups.

Table 8. Accuracy measures of those traditional fluency measures that significantly differed between the two groups based on the previous comparative statistic tests

Fluency tasks	Traditional measures	Accuracy measures					
		p	AUC	95% CI-	95% CI+	Sensitivity (%)	Specificity (%)
Letter 'a'	Number of errors	0.116	0.630	0.474	0.785	36.0	88.0
Animals	Number of correct words	**0.000**	0.842	0.734	0.949	100	56.0
Food items	Number of correct words	**0.002**	0.750	0.616	0.884	76.0	64.0
Actions	Number of correct words	**0.022**	0.689	0.543	0.834	68.0	52.0

AUC: area under the curve; CI: confidence interval.
Significant *p*-values (**$p < 0.05$**) indicate that the measure is significantly better than chance at discriminating individuals of the two groups.

Therefore, the described method would not require additional time-consuming steps, such as the manual transcription and preparation of the answers, nor their identification as correct words, errors, repetitions, or clusters, as opposed to the majority of fluency analysis techniques. This could make the analysis procedure considerably faster and easier. However, since this method does not provide any semantic information, it can be viewed for example as an alternative, inverse approach of the traditional analyses based on word count, because instead of considering the number of recalled words, this method focuses on the silent pauses between the words.

Our results confirmed the advantage of semantic fluency in the detection of MCI. In all three semantic fluency tests (animal, food item, and action), the same three temporal parameters (number of silent pauses, average length of silent pauses, average word transition time), and one of the traditional measures (correct word count) showed differences between the two groups. In contrast, regarding the phonemic fluency tests, differences were only observed in the case of the 'a' phonemic fluency, where two temporal parameters (the average and total length of irrelevant utterances) and one of the traditional measures (incorrect words) showed significant difference. These results are consistent with those of earlier studies, confirming that semantic fluency tasks may be

more appropriate for detecting the cognitive changes that occur in MCI (McDonnell et al., 2020; Nikolai et al., 2018). Furthermore, when compared to other subtypes of semantic fluency tests (plants, clothes, vehicles), the animal fluency test has previously shown the highest sensitivity (98.8%) in discriminating between HC and MCI participants (García-Herranz et al., 2020). In agreement with the results of García-Herranz et al., animal fluency achieved the best accuracy scores in the present study as well, not only with the traditional scoring method but also when examining the temporal parameters.

Limitations

The significant age difference between the HC and MCI groups may be noted as a limitation of this study, although elderly age itself is a primary risk factor of MCI. However, it has been also suggested that age has a significant influence on verbal fluency abilities (Kempler, Teng, Dick, Taussig, & Davis, 1998; Rodriguez-Aranda & Martinussen, 2006). Thus, we cannot rule out the possibility that the age of the participants might have affected their verbal fluency performance regardless of their cognitive state. Nevertheless, this

sample would closely represent the affected population in case of a potential real-life application.

When interpreting the results, it is important to take into consideration that because of the exploratory nature of this pilot study, corrections for multiple comparisons were not applied during the statistical analysis. As one of the main goals of this study was to investigate and identify all temporal fluency parameters that are able to differentiate between the groups, confirmatory studies are required to further attest the discriminatory ability and clinical utility of these significant temporal parameters.

This study established the main characteristics of a novel verbal fluency analysis, thus, further projects should be focused on the collection of more and higher quality data in order to define precise reference values for the amount of silent pauses associated with MCI. In the future, this would allow for the development of an automated tool for MCI screening, based on the analysis of temporal speech parameters. In addition, it remains to be determined whether combining this method of temporal parameter analysis with automated clustering analysis (reported earlier, e.g., König et al., 2018) could provide additional value with respect to classification.

CONCLUSION

In this study, we offered an alternative method of fluency analysis, and demonstrated the discriminatory ability of silent pause parameters in the case of semantic verbal fluency tests. Silence-related parameters can be extracted and calculated from fluency voice recordings using computerized methods. Therefore, this approach to fluency analysis seems to show promising potential, and, building on these results, the next step would be to construct an automated instrument capable of identifying MCI patients based on their speech/silence ratio. The development of remote, automated tools is especially important, seeing that the necessity and significance of medical consultations based on telemedicine are becoming common practice due to the current COVID-19 pandemic. Considering the high burden on the healthcare systems, an automated and cost-effective telemedical tool, based on the recognition of silent segments of speech, would be a valuable addition to practice, and it would likely improve the detection rates of MCI.

ACKNOWLEDGMENTS

The authors wish to thank all participants for their cooperation.

FINANCIAL SUPPORT

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

ETHICAL STANDARDS

Participation in the study was voluntary. All participants were informed about the aims of the study and gave their written consent. The experiment was conducted according to the ethical principles of the Declaration of Helsinki, and it was approved by the Regional Human Biomedical Research Ethics Committee of the University of Szeged (Reference No. 231/2017-SZTE).

REFERENCES

Boersma, P. & Weenink, D. (2020). *Praat: Doing Phonetics by Computer [Computer Program] (6.1.28) [Computer Software]*. Amsterdam, The Netherlands: Phonetic Sciences.

Borkowski, J.G., Benton, A.L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*, 135–140.

Chen, L., Asgari, M., Gale, R., Wild, K., Dodge, H., & Kaye, J. (2020). Improving the assessment of mild cognitive impairment in advanced age with a novel multi-feature automated speech and language analysis of verbal fluency. *Frontiers in Psychology*, *11*, 535.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, *44*, 837–845.

Demetriou, E. & Holtzer, R. (2017). Mild cognitive impairments moderate the effect of time on verbal fluency performance. *Journal of the International Neuropsychological Society*, *23*, 44–55.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Folstein, M.F., Folstein, S.E., & Mchugh, R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

García-Herranz, S., Díaz-Mardomingo, M.C., Venero, C., & Peraita, H. (2020). Accuracy of verbal fluency tests in the discrimination of mild cognitive impairment and probable Alzheimer's disease in older Spanish monolingual individuals. *Aging, Neuropsychology, and Cognition*, *27*, 826–840.

Gathercole, S.E., Willis, C.S., Baddeley, A.D., & Emslie, H. (1994). The children's test of nonword repetition: a test of phonological working memory. *Memory*, *2*, 103–127. https://doi.org/10.1080/09658219408258940

Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J., & Hoffmann, I. (2019). Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language*, *53*, 181–197.

Hahn, E.A. & Andel, R. (2011). Nonpharmacological therapies for behavioral and cognitive symptoms of mild cognitive impairment. *Journal of Aging and Health*, *23*, 1223–1245.

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*, 1212–1222.

Holmlund, T.B., Cheng, J., Foltz, P.W., Cohen, A.S., & Elvevåg, B. (2019). Updating verbal fluency analysis for the 21st century: applications for psychiatry. *Psychiatry Research*, *273*, 767–769.

IBM Corp. (2016). *SPSS Statistics for Windows (24.0) [Computer Software]*. Armonk, NY: IMB Corporation.

Jacobs, S., Mercuri, G., & Holtzer, R. (2021). Assessing within-task verbal fluency performance: the utility of individual time intervals in predicting incident mild cognitive impairment. *Aging, Neuropsychology, and Cognition*, *28*, 733–747.

Kempler, D., Teng, E.L., Dick, M., Taussig, I.M., & Davis, D.S. (1998). The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society*, *4*, 531–538.

Knopman, D.S. & Petersen, R.C. (2014). Mild cognitive impairment and mild dementia: a clinical perspective. *Mayo Clinic Proceedings*, *89*, 1452–1459.

König, A., Linz, N., Tröger, J., Wolters, M., Alexandersson, J., & Robert, P. (2018). Fully automatic speech-based analysis of the semantic verbal fluency task. *Dementia and Geriatric Cognitive Disorders*, *45*, 198–209.

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*, 205–220.

Lezak, M.D. (2012). *Neuropsychological Assessment* (5th ed.). New York: Oxford University Press.

Lopez-de-Ipina, K., Martinez-de-Lizarduy, U., Barroso, N., Ecay-Torres, M., Martinez-Lage, P., Torres, F., & Faundez-Zanuy, M. (2015). *Automatic analysis of categorical verbal fluency for mild cognitive impartment detection: a non-linear language independent approach. 2015 4th International Work Conference on Bioinspired Intelligence (IWOBI)*, San Sebastian, Spain, 101–104.

McDonnell, M., Dill, L., Panos, S., Amano, S., Brown, W., Giurgius, S., . . . Miller, K. (2020). Verbal fluency as a screening tool for mild cognitive impairment. *International Psychogeriatrics*, *32*, 1055–1062.

MedCalc Software Ltd (2020). *MedCalc Software (19.6.1.) [Computer Software]*. Ostend, Belgium: MedCalc Software Ltd.

Newcomb, F. (1969). *Missile Wounds of the Brain*. Oxford: Oxford University Press.

Nikolai, T., Bezdicek, O., Markova, H., Stepankova, H., Michalec, J., Kopecek, M., . . . Vyhnalek, M. (2018). Semantic verbal fluency impairment is detectable in patients with subjective cognitive decline. *Applied Neuropsychology: Adult*, *25*, 448–457.

Östberg, P., Fernaeus, S.-E., Hellström, Å., Bogdanović, N., & Wahlund, L.-O. (2005). Impaired verb fluency: a sign of mild cognitive impairment. *Brain and Language*, *95*, 273–279.

Pakhomov, S.V.S. & Hemmy, L.S. (2014). A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun study. *Cortex*, *55*, 97–106.

Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*, 183–194.

Piatt, A.L., Fields, J.A., Paolo, A.M., & Tröster, A.I. (1999). Action (verb naming) fluency as an executive function measure: convergent and divergent evidence of validity. *Neuropsychologia*, *37*, 1499–1503.

Roberts, R.O. & Knopman, D.S. (2013). Classification and epidemiology of MCI. *Clinics in Geriatric Medicine*, *29*, 753–772.

Roberts, R.O., Knopman, D.S., Mielke, M.M., Cha, R.H., Pankratz, V.S., Christianson, T.J.H., . . . Petersen, R.C. (2014). Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology*, *82*, 317–325.

Rodriguez-Aranda, C. & Martinussen, M. (2006). Age-related differences in performance of phonemic verbal fluency measured by controlled oral word association task (COWAT): a meta-analytic study. *Developmental Neuropsychology*, *30*, 697–717.

Rosen, W.G., Mohs, R.C., & Davis, K.L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, *141*, 1356–1364.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. Newbury Park: Sage.

Shao, Z., Janse, E., Visser, K., & Meyer, A.S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772.

Shulman, K.I., Shedletsky, R., & Silver, I.L. (1986). The challenge of time: clock-drawing and cognitive function in the elderly. *International Journal of Geriatric Psychiatry*, *1*, 135–140.

Sluis, R.A., Angus, D., Wiles, J., Back, A., Gibson, T., Liddle, J., . . . Angwin, A.J. (2020). An automated approach to examining pausing in the speech of people with dementia. *American Journal of Alzheimer's Disease & Other Dementias*, *35*, 153331752093977.

Taler, V., Johns, B.T., & Jones, M.N. (2020). A large-scale semantic analysis of verbal fluency across the aging spectrum: data from the Canadian longitudinal study on aging. *The Journal of Gerontology Series B: Psychological Science and Social Science*, *75*, e221–e230.

Teng, E., Leone-Friedman, J., Lee, G.J., Woo, S., Apostolova, L.G., Harrell, S., . . . Lu, P.H. (2013). Similar verbal fluency patterns in amnestic mild cognitive impairment and Alzheimer's disease. *Archives of Clinical Neuropsychology*, *28*, 400–410.

Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., . . . Kalman, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research*, *15*, 130–138.

Tröger, J., Linz, N., König, A., Robert, P., Alexandersson, J., Peter, J., & Kray, J. (2019). Exploitation versus exploration – computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. *Neuropsychologia*, *131*, 53–61.

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*, 138–146.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale – Revised*. New York: The Psychological Corporation.

Woods, D.L., Wyma, J.M., Herron, T.J., & Yund, E.W. (2016). Computerized analysis of verbal fluency: normative data and the effects of repeated testing, simulated malingering, and traumatic brain injury. *PLoS One*, *11*, e0166439.

Yesavage, J.A. & Sheikh, J.I. (1986). 9/Geriatric depression scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*, 165–173. https://doi.org/10.1300/J018v05n01_09