Taylor & Francis
Taylor & Francis Group

Clinical Linguistics & Phonetics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iclp20

Narrative recall in relapsing-remitting multiple sclerosis: A potentially useful speech task for detecting subtle cognitive changes

Veronika Svindt, Gábor Gosztolya & Tekla E. Gráczi

To cite this article: Veronika Svindt, Gábor Gosztolya & Tekla E. Gráczi (2023): Narrative recall in relapsing-remitting multiple sclerosis: A potentially useful speech task for detecting subtle cognitive changes, Clinical Linguistics & Phonetics, DOI: 10.1080/02699206.2023.2170830

To link to this article: https://doi.org/10.1080/02699206.2023.2170830

Published online: 30 Jan 2023.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=iclp20

CLINICAL LINGUISTICS & PHONETICS
https://doi.org/10.1080/02699206.2023.2170830

Narrative recall in relapsing-remitting multiple sclerosis: A potentially useful speech task for detecting subtle cognitive changes

Veronika Svindt[a], Gábor Gosztolya[b], and Tekla E. Gráczi[a]

[a]Research Centre for Linguistics, Eötvös Loránd Research Network, Budapest, Hungary; [b]Eötvös Lorand Research Network - University of Szeged, Research Group on Artificial Intelligence, Szeged, Hungary

ABSTRACT
Our research studied relapsing-remitting multiple sclerosis (RRMS). In half of the RRMS cases, mild cognitive difficulties are present, but often remain undetected despite their adverse effects on individuals' daily life. Detecting subtle cognitive alterations using speech analysis have rarely been implemented in MS research. We applied automatic speech recognition technology to devise a speech task with potential diagnostic value. Therefore, we used two narrative tasks adjusted for the neural and cognitive characteristics of RRMS; namely narrative recall and personal narrative. In addition to speech analysis, we examined the information processing speed, working memory, verbal fluency, and naming skills. Twenty-one participants with RRMS and 21 gender-, age-, and education-matched healthy controls took part in the study. All the participants with RRMS achieved a normal performance on Addenbrooke's Cognitive Examination. The following parameters of speech were measured: articulation and speech rate, the proportion, duration, frequency, and average length of silent and filled pauses. We found significant differences in the temporal parameters between groups and speech tasks. ROC analysis produced high classification accuracy for the narrative recall task (0.877 and 0.866), but low accuracy for the personal narrative task (0.617 and 0.592). The information processing speed affected the speech of the RRMS group but not that of the control group. The higher cognitive load of the narrative recall task may be the cause of significant changes in the speech of the RRMS group relative to the controls. Results suggest that narrative recall task may be effective for detecting subtle cognitive changes in RRMS.

ARTICLE HISTORY
Received 29 November 2021
Revised 6 January 2023
Accepted 11 January 2023

KEYWORDS
Relapsing-remitting multiple sclerosis (RRMS); narrative recall; personal narrative; automatic speech recognition technology (ASR); ROC analysis (receiver operating characteristics)

Introduction

Although complex cognitive decline is initially absent in multiple sclerosis (MS), especially in relapsing-remitting MS (RRMS), a significant number of patients still have difficulties in daily life due to the slowdown of the information processing speed, and decrease in directed attention and memory (e.g. working memory and episodic memory) (Rao, 2004; Sumowski et al., 2018).

In clinical practice, many methods and tasks have a diagnostic value for one or more particular disorders to detect cognitive difficulties. For example, verbal fluency (VF) tasks

CONTACT Veronika Svindt svindt.veronika@nytud.hu Research Centre for Linguistics, Eötvös Loránd Research Network, Budapest, Hungary
© 2023 Taylor & Francis Group, LLC

are generally used in neurodegenerative diseases. These tasks are a common part of English-speaking MS assessments because of their sensitivity to the cognitive difficulties in MS, especially to information processing efficiency (Beatty, 2002; Henry & Beatty, 2006; Lebkuecher et al., 2021). Some of the VF studies focus on the neural pathways activated by the fluency tasks (Blecher et al., 2019; Santiago et al., 2007; Woods et al., 2005). The results of Blecher et al. (2019) showed that the phonemic fluency task activated the dorsal language stream, and semantic fluency activated the ventral stream. According to Woods et al. (2005), the action fluency task effectively revealed impairments in the frontal – striatal–thalamocortical loops. Some parts of these loops are commonly affected in MS (see Table 1 on the common brain lesions in MS). Other studies suggested that action fluency frequently shows significant difficulties in MS because of the impairment of the input and output pathways of the prefrontal cortex (see Santiago et al., 2007). Although common cognitive tasks – such as VFs – are usually suitable for more than one neurological disease, not all tasks and methods are suitable for diagnosing or predicting all neural impairments due to the diverse neurological backgrounds of these disorders.

Language and speech are susceptible to subtle cognitive changes as several cognitive processes are simultaneously activated during speaking, and even mild difficulties in one of those processes can cause changes in speech production. The information processing speed and efficiency has a measurable impact on individuals' speech parameters (cf. Rodgers et al., 2013). Identifying the interactions between speech and other cognitive processes and describing the changes in speech and cognition may have diagnostic significance when managing and assessing neurodegenerative disorders. To date, only a few studies have investigated the changes of temporal parameters (henceforth: TPs) of speech in response to different speech tasks and cognitive load in MS (Feenaughty et al., 2013; Svindt et al., 2020). The type of speech task may affect acoustic, phonetic and grammatic parameters of speech in speakers with a neurodegenerative disease in various ways (Kempler & Van Lancker, 2002; Sidtis et al., 2015; Van Lancker Sidtis et al., 2012). Feenaughty et al. (2013) and Svindt

Table 1. Commonly affected brain regions and associated cognitive and narrative functions.

Common brain atrophy or lesions in (RR)MS	Associated cognitive function
Abnormal brain activity in the left dorsolateral PFC (Du et al., 2019)	• monitoring and manipulation of the content of working memory during narrative comprehension • casual-temporal ordering • casual-temporal orderingdifficulty in initiating, continuing or completing a story (Mar, 2004)
Abnormal brain activity in the left orbitofrontal PFC (Du et al., 2019)	• selecting appropriate details and inhibiting irrelevant elements during planning narrative comprehension and production (Mar, 2004)
Reduced grey matter and functional connectivity in the left ACC and PCC (Prinster et al., 2006; Rocca et al., 2012)	• initial learning, problem solving (Prinster et al., 2006) • autobiographic, episodic memory retrieval (Mar, 2004; Spreng et al., 2009) • narrative production (Mar, 2004)
Precuneus (Prinster et al., 2006)	• past episodes related to the self (Prinster et al., 2006)
White and grey matter loss in the cingulated-hippocampus network (Sacco et al., 2015)	• new learning, working memory (Sacco et al., 2015)
Left temporal pole (Steenwijk et al., 2016)	• narrative comprehension (Mar, 2004) • semantic processing
Temporoparietal junction (Sacco et al., 2015)	• episodic memory, prospection, navigation (Mar, 2004; Spreng et al., 2009)
Subcortical structures: caudate bilaterally (Prinster et al., 2006)	• associative learning and inhibitory control

PFC = prefrontal cortex; ACC = anterior cingulate cortex; PCC = posterior cingulate cortex.

et al. (2020) examined the effect of the cognitive demand of various speech tasks on several speech parameters. Despite the different methodologies, both studies found a strong effect of the more demanding speech tasks on the TPs of cognitively more impaired patients (Feenaughty et al., 2013), as well as progressive patients (Svindt et al., 2020).

As speech technology has advanced, it has become possible to accurately predict some cognitive changes from speech using automatic speech recognition (ASR) technology in various neurodegenerative diseases such as primary progressive aphasia, Parkinson's disease, mild cognitive impairment, Alzheimer's disease, and other dementias (e.g. Cuetos et al., 2007; Fraser et al., 2014; König et al., 2015; Lindsay et al., 2021; Meilán et al., 2012; Moro-Velazquez et al., 2019; Satt et al., 2014; Tóth et al., 2015, 2018). To the best of our knowledge, no study has analysed the TPs of speech in RRMS with ASR technology. The rationale for this approach over more traditional ones (such as the manual analysis by Praat) is that it is significantly less labour-intensive, yet it still provides an accurate evaluation of speech parameters.

The aim of our study is to apply two carefully selected speech tasks and analyse them with ASR to detect and predict initial cognitive decline. Therefore we only investigated the relapsing-remitting type of MS in the remitting period and only those participants who fell within the normal range on neuropsychological assessment. This study is the first step towards creating a reliable screening tool for clinicians.

Relapsing-remitting multiple sclerosis and cognition

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system associated with demyelinating processes. The neurobiological consequence of myelin loss is a significant deceleration in the information transmission efficiency along the axons. Because of this, one of the first main symptoms is the slowdown of the information processing speed (Rao, 2004; Sumowski et al., 2018), which is likely to affect other cognitive processes (e.g. speech).

The present study focuses on the relapsing-remitting MS type, which occurs in 80% of all MS cases. In RRMS, unpredictable relapsing periods with moderate-to-severe symptoms are followed by remitting periods without symptoms or only with mild symptoms. Most RRMS cases progress over time to secondary progressive MS (SPMS). The progression course is usually not apparently noticeable, but it would be helpful to accurately predict it.

Multiple sclerosis is associated with brain atrophy in various locations of the central nervous system (see Table 1), resulting in decreased neural network efficiency (Rocca et al., 2012). Although MS is primarily a white matter disease, several studies have shown that the brain volume (both white and grey matter), brain activation patterns, and functional connectivity between distinct brain areas deviate from healthy people in typical brain locations (Du et al., 2019; Prinster et al., 2006; Rocca et al., 2012; Sacco et al., 2015; Santiago et al., 2007; Sumowski et al., 2018).

An increasing number of studies indicate that cognition is often affected in RRMS (Prakash et al., 2008), and patients have mild-to-moderate cognitive difficulties in at least one cognitive domain early after the onset of the first symptoms. Commonly affected functions are information processing (general slowness in planning and execution), selective and focused attention, response inhibition, verbal and visuospatial analysis, long-term memory, especially episodic memory (encoding, storage, recall), verbal fluency and working memory (Chiaravalloti & De Luca, 2008; Matias-Guiu et al., 2020; Migliore et al., 2017; Nocentini et al., 2006; Olivares et al., 2005; Prakash et al., 2008; Rocca et al., 2012; Sumowski et al., 2018).

Aims and hypotheses

We sought to create an appropriate method to reveal any subtle impairments that would remain undiscovered in RRMS with general non-MS-specific neuropsychological assessments. Thus, we only included those RRMS participants in the remission phase who had normal performance on the Hungarian version of the Addenbrooke's Cognitive Examination (ACE, Stachó et al., 2003).

During the selection of the speech tasks, we focused on the frequently impaired cognitive functions (see above) and the typically affected brain structures in RRMS and their connection with narrative comprehension and production (Table 1). Therefore, we chose two narrative speech tasks that activate neural and cognitive processes likely to be affected in RRMS. These are:

(1) A personal narrative about the past (events of their previous day);
(2) Narrative (story) recall task.

Task (1) mainly activates autobiographical memory. Former studies of our colleagues indicated that an ASR analysis of this type of personal narrative seemed to have a diagnostic value in the detection of mild cognitive impairment (Tóth et al., 2015, 2018). Task (2) is one of the most challenging speech tasks not only for people suffering from neurodegenerative diseases but also for healthy adults (Bóna, 2014). Although participants have to produce coherent, complex narratives in both tasks, there are some significant differences in the cognitive requirements of these tasks (Table 2). We hypothesised that both narrative tasks might be effective to measure some differences in the TPs of speech between RRMS participants and healthy controls because both tasks require the activation of commonly affected brain areas in RRMS. Moreover, we predicted more significant differences in the narrative recall relative to the personal narrative between groups due to the higher cognitive complexity of the recall task. Speech tasks are evaluated with the ASR technology presented below.

We also explored specific cognitive abilities likely to be affected in RRMS despite normal performance on ACE. Hence, we measured information processing speed, working memory, verbal fluency, naming and sentence repetition skills and the effect and/or interaction of these features on the speech in RRMS.

Table 2. Main cognitive requirements of the tasks.

Task 1 : Personal narrative about the past	Task 2: Narrative (story) recall
Narrative production	Narrative comprehension + production
Personal and emotional	Non-personal and non-emotional
Episodic (autobiographical) memory	Episodic buffer (working memory) (see: Baddeley & Wilson, 2002)
Storage + retrieval	Encoding + storage + (immediate) retrieval
Inhibition of irrelevant events	Inhibition of irrelevant information
Sequencing and organising (events, space and time)	
Sustained, focused attention	

Methods

Participants

A total of 42 people participated in the study: 21 people diagnosed with RRMS and 21 age- and gender- and education-matched healthy controls (HC) (Table 3). Our matching procedure occurred at the group level; one-way ANOVA (for age and education) and Chi-square test (for gender) did not reveal significant differences between groups. All the participants were native Hungarian speakers with intact hearing and without any depressive disorder (measured by the Beck's Depression Inventory, BDI-II, Beck et al., 1996) or other known psychiatric conditions. An average performance (at least 90 points) on the ACE was the main inclusion criterion for the RRMS participants. We expected an average performance on ACE in the case of controls. RRMS participants had mild-to-moderate physical disability according to the Expanded Disability Status Scale (EDSS, Kurtzke, 1983). All the participants were volunteers and gave their informed consent. The Ethics Committee of the Uzsoki Hospital, Budapest approved the study.

Procedure

The procedure consisted of three separate sessions.

(1) A neuropsychologist and a neurologist evaluated the neuropsychological (ACE) and physical (EDSS) examination in the first session.
(2) In the second session, we recorded the narrative speech tasks and conducted cognitive tasks. Each participant performed two speech tasks: (2a) a personal narrative about the past *('Tell me about your yesterday in as much detail as possible from morning to evening!'*) and (2b) a narrative (story) recall task. For the narrative recall task, we used a two-minute-long educational text presented from a recording. The text was about the subconscious role of touch in our everyday life and did not contain difficult-to-understand or scientific terms. A trained speech expert read it for the recording. After listening to the recording, participants had to summarise the content of the text as accurately as possible.

Then, in the same session, we used working memory tasks (digit span forward and backwards; non-word repetition, reading span, and sentence repetition); verbal fluency task; and the Boston Naming Task (BNT) (Table 4). The tasks were conducted in a constant order, and the session lasted 40–45 minutes. Speech samples were recorded digitally with a Sony A-10 dictaphone and a clip-on microphone at 44.1 kHz sampling rate.

Table 3. Participants.

	RRMS ($n = 21$)	HC ($n = 21$)
Age (in years, mean)	39.0 [range: 24–56]	40.1 [range: 28–56]
Male : female	5 : 16	5 : 16
Education (in years, mean)	15,1 [range: 12-19]	16,0 [range: 12-19]
Post onset (in years, mean)	9 [range: 1–20]	–
EDSS score (mean)	3.2 [range: 0,0–6,5]	–
ACE score (mean)	96 [range: 91–100]	–

Table 4. Cognitive tasks.

Measured cognitive ability	Task	Standard reference values
Executive functions (here: cognitive flexibility and information processing speed)	Wisconsin Card Sorting Task (WCST, Grant & Berg, 1948)	No (groups were compared to each other)
	Stroop task (Stroop, 1935)	
Working memory	Digit span task (forward)	International standard (Grégoire & Van der Linden, 1997)
	Digit span task (backwards)	
	Non-word repetition	Hungarian standard (Racsmány et al., 2005)
	Reading span task	Hungarian standard (Racsmány et al., 2005)
	Sentence repetition task	The task is a part of the Hungarian working memory battery (Racsmány et al., 2005), but it has no reference values
Verbal fluency	Phonemic fluency (letter <m>)	ACE international reference values
	Semantic fluency (aminals)	
	Action (verb) fluency	No (groups were compared to each other)
Naming	Boston Naming Task (BNT)	International standard (Kaplan et al., 1983)

(3) The final session was an online test set consisting of the Wisconsin Card Sorting Task (WCST) and the Stroop task. This online session was completed in the participants' homes. We used an online interface based on the open-source Psytoolkit platform for the WCST and Stroop tasks (https://www.psytoolkit.org/). Participants were instructed to complete the tasks (a) on a computer or a laptop; (b) with a mouse. In addition, we asked them to (c) do the tasks without interruption, (d) not do the task if they felt significant fatigue, and (e) do it within one week after the speech recording. Five participants (4 RRMSs, 1 HC) did not complete the online tasks. We made a special effort to exclude all technical factors that might affect any parameters of the online test interface.[1]

Analysis of the cognitive tasks

We measured cognitive flexibility and information processing speed with the WCST and Stroop tasks. In these tasks, total errors, perseveration errors and reaction time were measured. Digit span (forward and backwards), non-word repetition, reading span, and sentence repetition tasks were used for evaluating working memory capacity. These tasks are part of the Hungarian working memory test battery (Racsmány et al., 2005). In the sentence repetition task, participants repeated 10 sentences, similar in syntactic complexity and length (13–16 words per sentence). We also measured verbal fluency skills with three VF tasks: phonemic, semantic and action fluency. Naming ability was measured with the Boston Naming Task. The comparison between groups was based on the international or Hungarian standard values or – in absence of standards – on the performance of the groups (see Table 4). International standard values were used in the case of the two digit span tasks, the phonemic and semantic fluency task, and in the BNT. Hungarian standard values were used for the non-word repetition and the reading span tasks. The online version of the WCST and Stroop task, the sentence repetition task, and the action fluency task had no reference values. Therefore, we compared the results of the RRMS group with the control group in these tasks.

[1]We checked the online tests using the following technical factors. Software: the test ran on all operating systems. Browser: if an outdated browser was detected, the user was instructed to instal an updated one. Hardware: very low computing power was required for the tasks. Monitor size and display resolution: the tasks were adapted to these factors.

Analysis of the narrative speech tasks

The two narrative speech samples (personal narrative and narrative recall) were analysed using automatic speech recognition (ASR) technology. Our ASR system used standard speech recognition techniques. We employed the HTK tool (Young et al., 2006), which was modified to allow the use of a Hidden Markov Model/Deep Neural Network (DNN) hybrid set-up (Hinton et al., 2012). We used 40 raw Mel-frequency filter bank energy values as acoustic features along with the global log-energy, which was extended with the first and second order derivatives ('FBANK + Δ + ΔΔ'), resulting in 123 acoustic features overall. Training and evaluation were performed on a 150 ms (15 frames) wide sliding window (resulting in 1845 input neurons). Next, the DNN contained 5 fully connected hidden layers, each consisting of 1024 ReLU neurons (Glorot et al., 2011), while the final layer had 911 neurons (i.e. equal to the number of phonetic states) with the softmax activation. The DNN acoustic model was trained on a subset of roughly 60 hours of recordings from the BEA corpus (Neuberger et al., 2014).

Our ASR model produced a time-aligned token sequence for each recording; that is, it supplied a hypothesis of the sequence of tokens uttered, along with the starting and ending time indices. The set of tokens consisted of phones of the Hungarian language along with special tokens for silent and filled pauses. Therefore, we distinguished two types of tokens, namely phoneme tokens and pause tokens. Pause tokens consisted of silent and filled pauses. The latter included hesitations (*um, er*, etc.) and lengthenings. From the output of this ASR system, we calculated the following parameters:

(1) Speech rate (the number of syllables per second including hesitations in the total duration of the speech sample);
(2) Articulation rate (the number of syllables per second excluding hesitations during the speech);
(3) Pause duration ratio (the duration of pause tokens divided by the length of the speech sample);
(4) Pauses count ratio (the number of pause tokens divided by the overall number of tokens)
(5) Pause frequency (the number of pause tokens divided by the length of the speech sample);
(6) Average pause duration (the total duration of pause tokens divided by the number of pauses).

Statistical analysis

The statistical analyses were carried out in R (R Core Team, 2021).

Multiple linear regression models (lme4, Bates et al., 2015.) were run to test whether (i) the cognitive tasks (dependent variable) were affected by the post-onset time or the EDSS score (factors) in the MS patients and by (ii) the age and speaker group (factors, interaction allowed). The p-values were obtained by the Satterthwaite-approximation (anova()), and the adjusted determination coefficients were obtained by summary(). The age did not have any significant effect, so (ii) was simplified to the Mann – Whitney test. One extreme outlier

(12470.29 ms) appeared in the WCST test's mean reaction times that must have appeared due to technical factors or the participant must have left the test at a point and returned to it several minutes later. This data was excluded from the statistics of the intergroup differences of WCST.

Linear mixed models (henceforth: lmms) (lme4) were run to test whether the temporal parameters (TPs) (dependent variable) were influenced by (i) the group and speech task (fixed factors with interaction), (ii) the Stroop task reaction time (RT), WCST RT and the speech task (fixed factors with interaction), (iii) the EDSS score and post-onset time (fixed factors, no interaction) within the speech tasks (fixed factor, with possible interaction with the EDSS score and the post-onset time). The results of both groups were included in (i), while (ii) was evaluated separately. In (iii) only the MS group was tested. The p-values were obtained by the Satterthwaite-approximation (anova()). Model selection was only carried out for (i) and (iii) (lmerTest, Kuznetsova et al., 2017). No model selection was carried out for (ii), where only the effect of Stroop RT was examined. The lmms included only random intercepts by the speakers. Effect sizes were calculated (MuMIn: Bartoń, 2022, afex: Singmann 2021).

Results

Cognitive tasks

First, we converted the raw data of the RRMS group in the cognitive tasks into z-scores using Hungarian or international standard reference values. In absence of standard values, z-scores of the RRMS group were calculated using the mean scores and SD of the control group (for the comparison method of each task, see Table 4; for the results in z-scores, see Figure 1). Second, we compared the raw scores of the groups with the Mann-Whitney test (Table 5).

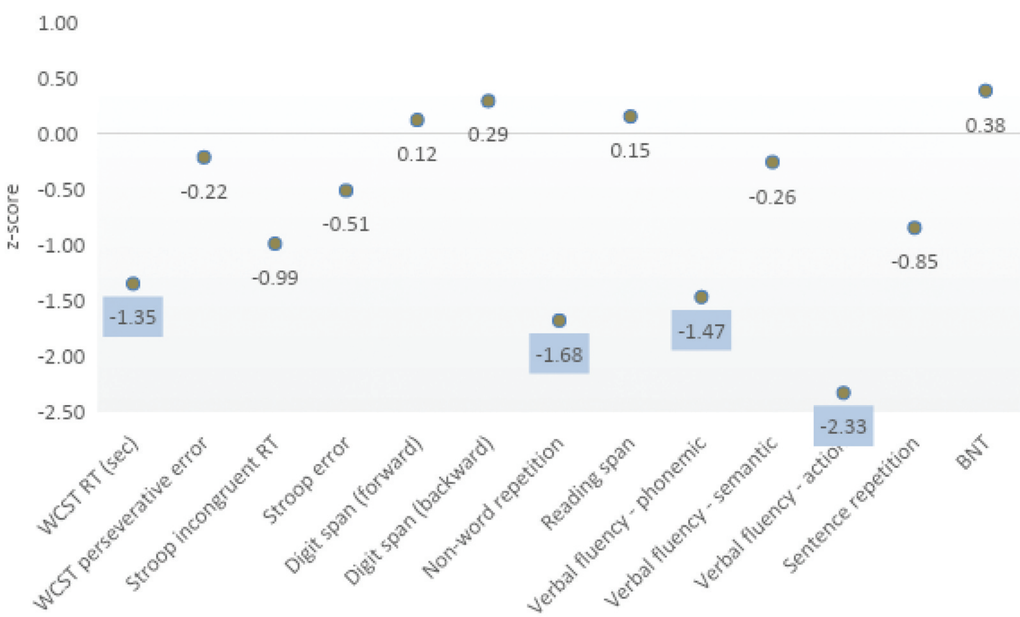

Figure 1. Results of the cognitive tasks in the RRMS group compared to standard test values (z-scores).

Table 5. Descriptives and the Mann-Whitney test of the cognitive tasks.

Cognitive ability	Task	RRMS [*n* = 21]	HC [*n* = 21]	p
Executive functions (here: cognitive flexibility and information processing speed)	WCST			
	• perseverative error (mean)	5.00	4.35	.604
	• RT (ms, mean, SD)	3678.23 ± 1122.87	3080.85 ± 859.59	.089
	Stroop task			
	• error	0.47	0.20	.311
	• congruent items RT (ms, SD)	2283.79 ± 637.79	1878.24 ± 374.13	.033
	• incongruent items RT (ms, SD)	2068.86 ± 548.37	1771.55 ± 292.87	.110
Working memory	Digit span task (forward)	6.21 ± 1.08	6.68 ± 0.95	.201
	Digit span task (backward)	5.11 ± 1.24	5.72 ± 1.27	.159
	Non-word repetition	5.61 ± 0.85	6.84 ± 0.50	<.001
	Reading span task	3.89 ± 1.49	4.94 ± 1.11	.034
	Sentence repetition task	2.1 ± 2.46	4.0 ± 2.19	.005
Verbal fluency	Phonemic fluency	16.11 ± 4.06	20.63 ± 4.54	<.001
	Semantic fluency	25.57 ± 4.85	33.24 ± 5.34	<.001
	Action fluency	27.57 ± 6.92	38.19 ± 4.17	<.001
Naming	Boston Naming Task (BNT) raw score	57.1 ± 3.36	58.7 ± 1.35	.077

Both groups performed similarly in the WCST and the Stroop task. However, the Mann-Whitney test revealed significant differences between groups in the reaction time of the Stroop congruent items: the RRMS group was significantly slower than the controls. Working memory tasks showed a different picture. In the digit span tasks, the performance of the RRMS group did not differ significantly from the international standard values or the control group. Non-word repetition, reading span, and sentence repetition tasks showed significant differences between groups (Table 5). Relative to the age-matched standard scores, RRMS participants had significantly lower performance in the non-word repetition task, but not in the reading span task (Figure 1). All three verbal fluency tasks displayed significant differences between groups (Table 5). However, comparing the phonemic and semantic fluency results to ACE standards, only the phonemic fluency task displayed a negative difference relative to controls. There is no Hungarian standard for action fluency task; the comparison between groups demonstrated a significant difference: RRMS participants listed an average of 11 fewer verbs in 1 minute than controls did. Most RRMS participants had no word-finding difficulties according to BNT scores, and only one person had a below-average performance on this task.

According to the linear regression models, neither the post-onset time nor the EDSS scores significantly affected any of the information processing tasks (WCST and Stroop). The effect of these factors was attested to the TPs in lmms with the possible interaction of speech task. The model selection gave different models as best fits for the TPs. However, none of the factors had any significant effect on the TPs.

ASR analysis of the narrative speech tasks

Figure 2 shows the automatic speech recognition analysis, i.e. the temporal parameters for both narrative speech tasks. Table 6 lists the linear mixed model results for these parameters by testing the possible effect of the factors *speech task* and *speaker group*. The best-fitting model based on the model selection and the factor or interaction that had a significant effect on the TPs is shown in this table. The interaction of the two factors had a significant effect

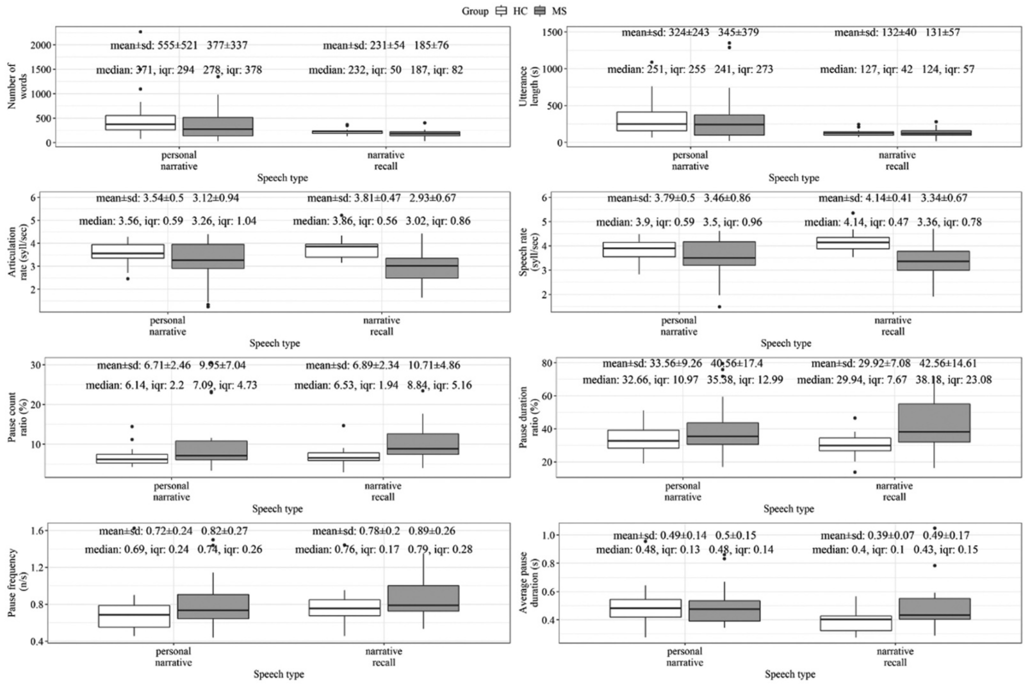

Figure 2. Temporal parameters of the speech tasks.

Table 6. The results of the LMMs for the factors group and narrative speech task on the TPs (NumDF = 1 in each case).

Variable	Best model	Sign. effect	DenDF	F	p	r^2_m	r^2_c
Number of words	Only speech task	Speech task	41	14.413	<.001	0.147	0.174
Average length of the narrative (sec)	Only speech task	Speech task	41	19.31	<.001	0.173	0.274
Articulation rate (syll)	Speech task*group	Speech task*group	40	4.918	.032	0.221	0.608
Speech rate (syll)	Speech task*group	Speech task*group	40	5.852	.020	0.207	0.591
Pauses count ratio (pause tokens/ overall tokens)	Group	Group	40	7.420	.010	0.138	0.716
Pauses duration ratio (sec/sec)	Speech task*group	Group	40	7.538	.009	0.146	0.700
Pauses frequency ratio (pause tokens/ sec)	Speech task	Speech task	41	7.394	.009	0.020	0.782
Pauses average length (sec/pause tokens)	Speech task*group	Speech task*group	40	6.048	.018	0.096	0.684

on the articulation and speech rates, and the average pause length. The Tukey post hoc test revealed a significant difference in the articulation rate and the speech rate between groups in the narrative recall tasks (with slower rates for the RRMS group), and a significant difference in the average pause length between the two speech tasks within the HC group. The number of words, the pause count ratio and the pause duration ratio were significantly different between the HC and MS groups regardless of the speech task (with higher values for the latter), and the average utterance length and the pause frequency were found to be significantly different between the speech tasks regardless of the speaker groups (with lower values and variability in the narrative recall). With most of the measures, the variability was lower in the HC group. The effect sizes showed large differences in a few models (see Table 6), and indicated that the random intercept by speakers increased considerably the model's explanatory power in these cases perhaps due to the larger variability in the MS group.

The effect of the information processing efficiency on the speech parameters

Although the Mann-Whitney test did not reveal significant differences in the RT of the Stroop incongruent items RT between groups, these variables may tell us something about the difficulties in the information processing efficiency. Therefore, we decided to evaluate the relation between the TPs and the RT in the Stroop and WSCT tasks. The Stroop RT did not have any significant effect in the HC group on any of the TPs regardless of the speech task, but it did on most TPs in the RRMS group in both speech tasks (see Table 7). The large differences between the marginal and conditional effect sizes indicate that the random intercept improved the model fit considerably, perhaps due to the large individual differences already seen in the group – speech task comparisons.

As presented in the Introduction, verbal fluency tasks – especially phonemic and semantic fluency – are commonly used in clinical practice to measure the information processing efficiency in MS (e.g. Beatty, 2002; Henry & Beatty, 2006; Lebkuecher et al., 2021). Therefore we assessed correlation (between fluency tasks and information processing tasks) and analysed the possible effect of fluency tasks on TPs. In spite of the significant weak to moderate negative correlation between the semantic fluency and the Stroop incongruent RT in both groups (RRMS: $r = -.540$, $r < 0.001$; HC: $r = -.359$, $r = 0.023$), the semantic fluency performance did not influenced the TPs in the RRMS group. Phonemic fluency correlated only with the WCST RT in the clinical group ($r = -0.608$, $r = 0.016$), and had no effects on the TPs of the speech. The action fluency task had no correlation with any of the information processing tasks and had no effect on the TPs in any group.

Receiver operating characteristic (ROC)

We performed a receiver operating characteristic (ROC) analysis in order to evaluate the discriminative value of the two narrative speech tasks (method and reference values: Mandrekar, 2010). ROC curves for the articulation rate and speech rate attributes for the tasks were calculated by showing the true positive rate (sensitivity) against the false positive

Table 7. LMM results for the main effect, Stroop and the interaction WCST*speechtype on TPs in the RRMS group (DenDF: 1, NumDF: 20).

	Stroop task (main effect)				WCST*speech type interaction			
Variable	F	p	r2m	r2c	F	p	r2m	r2c
Speech rate	5.698	.031	0.226	0.570	4.932	.045	0.120	0.690
Articulation rate	5.941	.028	0.233	0.589	-	-	-	-
Pause count ratio	-	-	-	-	-	-	-	-
Pause duration ratio	4.838	.044	0.209	0.648	-	-	-	-
Pause frequency	6.339	.024	0.036	0.761	-	-	-	-
Pauses average length	6.339	.024	0.259	0.693	-	-	-	-

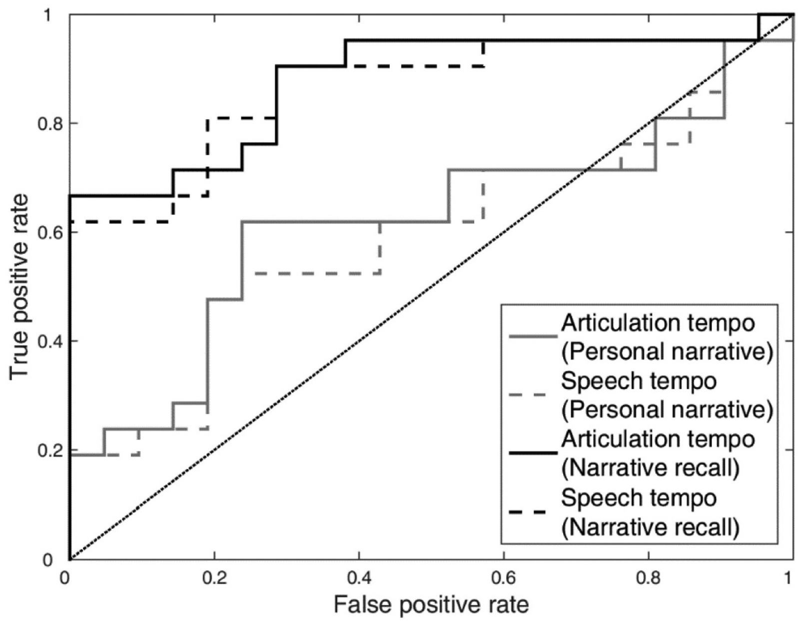

Figure 3. Receiver operating characteristic (ROC).

rate (1 - specificity) at each possible threshold setting (Figure 3), which is summarised with the area under the curve (AUC) value. All four curves exceed the dotted line (representing random guessing and having an AUC score of 0.5). For the personal narrative, the two corresponding curves are located lower; and they had no discriminatory ability (AUC values of 0.617 and 0.592, articulation rate and speech rate, respectively), while for narrative recall, they had excellent discriminatory ability (AUC values of 0.877 and 0.866, articulation rate and speech rate, respectively).

Discussion

Our initial hypothesis was that even early and subtle cognitive difficulties can be detected in (RR)MS via speech analysis, if the speech tasks were adjusted to the neural characteristics of the disease. We measured temporal parameters of speech with ASR technology, which proved to be suitable for a fast but reliable evaluation. Our results revealed (1) effects of the clinical group (RRMS group vs HC group); (2) effects of the speech task; (3) combined effects of (1) and (2) on the temporal parameters of the participants' speech; and (4) effect of the information processing efficiency on the speech parameters. Neither the disease duration nor the physical state (measured by the EDSS) affected the temporal parameters of speech and the performance on the executive function tasks (Stroop and WCST).

The effect of the group was shown in the pause count and pause duration ratios, regardless of the speech task. Persons with RRMS tend to speak with more and longer pauses, which may reflect the general deceleration of their information processing, and the consequence of the tasks' cognitive demand. The group effect was also shown in the performance on the cognitive tasks. Although none of our RRMS participants had any cognitive impairment according to ACE, significant deterioration was noted in verbal fluency and most working memory tasks (non-word repetition, reading span and sentence repetition) in the RRMS group.

However, despite the significant difficulties of the RRMS group in the verbal fluency tasks relative to the HC group, we found no effect of either task on the TPs in either group. In light of the studies on verbal fluency presented in the Introduction, the lack of this effect is somewhat unexpected. The narrative recall task presumably activated both dorsal and ventral streams and the loops seen in Table 1, and we showed the effect of this speech task on the RRMS group. These results indicate that other possible factors may have a significant effect on the temporal parameters of speech, which, however, cannot be identified based on the data collected in this study.

Furthermore, we observed difficulties in those working memory tasks which required the activation of the episodic buffer among others (reading span, sentence repetition). The episodic buffer needed to be active in the narrative recall task, but to a lesser extent or not at all in the personal narrative. Among the working memory tasks, the non-word repetition task showed the greatest difficulties in the RRMS group. Although the significant deterioration of the clinical group's performance could be caused by several reasons – e.g phoneme discrimination, subtle hearing or oralmotor problems –, this limitation together with the episodic buffer difficulties may have affected the results in the narrative recall task.

Our results also showed the effect of the speech task on the temporal parameters. Speech and articulation rate and the average length of pauses differed significantly only in the narrative recall task between groups. Furthermore, we found the combined effect of the speech task and the group in the narrative recall task, which revealed significant changes in the speech parameters only of the RRMS participants. ROC analysis produced excellent diagnostic accuracy (88% for the speech rate and 87% for the articulation rate) in the narrative recall task. In contrast, we did not get similar results for the personal narrative task which had no categorisation power for the RRMS group (62% and 59% for the speech and articulation rate). Although both narrative tasks were chosen for the typical neural impairments of the (RR)MS, the differences could be due to the considerable differences between the tasks. Retrieving familiar and/or personal events requires less cognitive load (see Abraham et al., 2008) than processing, storing (in the working memory) and retrieving unfamiliar and/or non-personal information. In addition, the personal narrative task was a speech production task, while the narrative recall task required both comprehension and production and overloaded the working memory, especially the episodic buffer. These features are closely related to the less efficient and slower information processing of the clinical group. Results showed the significant effect of the information processing efficiency (measured by the Stroop task's RT) on the temporal parameters in the RRMS but not in the HC group. This finding is in line with those studies that found a measurable effect of task complexity, task type, and/or the information processing speed on people's performance and/or speech production in MS (e.g. Denney et al., 2005 and 2011; Feenaughty et al., 2013; Rodgers et al., 2013; Svindt et al., 2020). According to Denney et al. (2005; 2011), the general slowness of the information processing in MS is relatively independent from other cognitive factors, but has a significant impact on them. Rodgers et al. (2013) demonstrated the predictive effect of the information processing speed on the articulation rate in MS. The finding in our current study of the effect of the speech task on the TPs in cognitively unimpaired RRMS patients converges with these earlier results.

Limitations

Our study has several limitations. First, due to the limited number of participants, only cautious conclusions can be drawn. Second, although none of our patients had marked dysarthria, some subtle motor speech disorders might have influenced our results. However, if differences between speech tasks in the RRMS group are partly caused by very mild dysarthria, this supports our hypothesis about the effect of the higher cognitive load on the RRMS group relative to the HCs. Finally, as the present study is only the first step of longitudinal research, the potentially predictive nature of the narrative recall task can be only confirmed or refuted in a future study.

Clinical implications and conclusion

Our results point to the cumulative effect of the greater cognitive demand combined with the less effective information processing on the speech of the RRMS patients. The narrative recall task showed excellent diagnostic accuracy in the RRMS vs. control differentiation. However, the personal narrative task was ineffective in measuring subtle cognitive changes in RRMS, in divergence from what was found earlier in the case of another neurodegenerative disease (namely in mild cognitive impairment, MCI). A previous study by Tóth et al. (2015, 2018) used the same personal narrative task and the same ASR technique for detecting mild cognitive impairment. Their results with the personal narrative task showed 75–90% diagnostic accuracy for the MCI group. The contrast between the categorisation power of the two narrative speech tasks in the same clinical population (RRMS) and the contrast between diagnostic groups (RRMS vs. MCI) in the same task (personal narrative) highlights the importance of the task selection method. Adapting the speech task to the neural characteristics of a particular clinical group appears to be crucial for potential diagnostic power.

Acknowledgements

We express our special thanks to professor Gábor Jakab, the chief physician of the Department of Neurology (Uzsoki Hospital, Budapest) and Rita Oláh, the tireless neuropsychologist of the Department. We also gratefully acknowledge the support of the Hungarian National Research, Development and Innovation Office (first and second authors: grant no. NKFI-K-132460; third author: grant no. NKFI-FK-128814). Special thanks also to our Reviewers for their valuable comments and supportive suggestions, which helped us to improve the quality of our manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Research, Development and Innovation Office [NKFI-K-132460,NKFI-FK-128814].

ORCID

Veronika Svindt http://orcid.org/0000-0002-6027-9029
Gábor Gosztolya http://orcid.org/0000-0002-2864-6466
Tekla E. Gráczi http://orcid.org/0000-0003-3351-9661

References

Abraham, A., Schubotz, R. I., & von Cramon, D. Y. (2008). Thinking about the future versus the past in personal and non-personal contexts. *Brain Research*, *1233*, 106–119. https://doi.org/10.1016/j.brainres.2008.07.084

Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, *40*(10), 1737–1743. https://doi.org/10.1016/S0028-3932(01)00146-4

Bartoń, K. (2022). MuMIn: Multi-Model Inference. R package version 1.46.0. https://cran.r-project.org/package=MuMIn

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beatty, W. (2002). Fluency in multiple sclerosis: Which measure is best? *Multiple Sclerosis*, *8*(3), 261–264. https://doi.org/10.1191/1352458502ms799oa

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory-II. *APA-Psych Tests*. https://doi.org/10.1037/t00742-000

Blecher, T., Miron, S., Schneider, G. G., Archiron, A., & Ben-Shachar, M. (2019). Association between white matter microstructure and verbal fluency in patients with multiple sclerosis. *Frontiers in Psychology*, *10*. Article 1607. https://doi.org/10.3389/fpsyg.2019.01607

Bóna, J. (2014). Temporal characteristics of speech: The effect of age and speech style. *The Journal of the Acoustical Society of America*, *136*(2), EL116–121. https://doi.org/10.1121/1.4885482

Chiaravalloti, N., & De Luca, J. (2008). Cognitive impairment in multiple sclerosis. *Lancet neurology*, *7*(12), 1139–1151. https://doi.org/10.1016/S1474-4422(08)70259-X

Cuetos, F., Arango Lasprilla, J. C., Uribe, C., Valencia, C., & Lopera, F. (2007). Linguistic changes in verbal expression: A preclinical marker of Alzheimer's disease. *Journal of the International Neuropsychological Society*, *13*(3), 433–439. https://doi.org/10.1017/S1355617707070609

Denney, D. R., Gallagher, K. S., & Lynch, S. G. (2011). Deficits in processing speed in patients with multiple sclerosis: Evidence from explicit and covert measures. *Archives of Clinical Neuropsychology*, *26*(2), 110–119. https://doi.org/10.1093/arclin/acq104

Denney, D. R., Sworowski, L. A., & Lynch, S. G. (2005). Cognitive impairment in three subtypes of multiple sclerosis. *Archives of Clinical Neuropsychology*, *20*(8), 967–981. https://doi.org/10.1016/j.acn.2005.04.012

Du, X. F., Liu, J., Hua, Q. F., & Wu, Y. J. (2019). Relapsing-remitting multiple sclerosis is associated with regional brain activity deficits in motor- and cognitive-related brain areas. *Frontiers in neurology*, *10*(November). https://doi.org/10.3389/fneur.2019.01136

Feenaughty, L., Tjaden, K., Benedict, R. H. B., & Weinstock-Guttman, B. (2013). Speech and pause characteristics in multiple sclerosis: A preliminary study of speakers with high and low neuropsychological test performance. *Clinical Linguistics & Phonetics*, *27*(2), 134–151. https://doi.org/10.3109/02699206.2012.751624

Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., & Rochon, E. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex*, *55*(1), 43–60. https://doi.org/10.1016/j.cortex.2012.12.006

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, PMLR* 15 (pp. 315–323).

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Grégoire, J., & Van der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition*, *4*(2), 140–149. https://doi.org/10.1080/13825589708256642

Henry, J. D., & Beatty, W. W. (2006). Verbal fluency deficits in multiple sclerosis. *Neuropsychologia*, *44*(7), 1166–1174. https://doi.org/10.1016/j.neuropsychologia.2005.10.006

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE signal processing magazine*, *29*(6), 82–97. https://doi.org/10.1109/MSP.2012.2205597

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston naming test*. Lea and Febiger.

Kempler, D., & Van Lancker, D. (2002). Effect of speech task on intelligibility in dysarthria: A case study of Parkinson's disease. *Brain and Language*, *80*(3), 449–464. https://doi.org/10.1006/brln.2001.2602

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P. H., & David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *1*(1), 112–124. https://doi.org/10.1016/j.dadm.2014.11.012

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, *33*(11), 1444–1452. https://doi.org/10.1212/WNL.33.11.1444

Kuznetsova, A., Brockhoff, B. F., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lebkuecher, A. L., Chiaravalloti, N. D., & Strober, L. B. (2021). The role of language ability in verbal fluency of individuals with multiple sclerosis. *Multiple Sclerosis and Related Disorders*, *50*, 102846. https://doi.org/10.1016/j.msard.2021.102846

Lindsay, H., Tröger, J., & König, A. (2021). Language impairment in Alzheimer's disease – Robust and explainable evidence for AD-related deterioration of spontaneous speech through multilingual machine learning. *Frontiers in aging neuroscience*, *13*. Article 642033. https://doi.org/10.3389/fnagi.2021.642033

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414–1434. https://doi.org/10.1016/j.neuropsychologia.2003.12.016

Matias-Guiu, J. A., Cortés-Martínez, A., Curiel, R. E., Delgado-Álvarez, A., Fernández-Oliveira, A., Pytel, V., Montero, P., Moreno-Ramos, T., Loewenstein, D. A., & Matías-Guiu, J. (2020). Memory impairment in relapsing-remitting multiple sclerosis using a challenging semantic interference task. *Frontiers in neurology*, *11*(April), 1–11. https://doi.org/10.3389/fneur.2020.00309

Meilán, J. J. G., Martínez-Sánchez, F., Carro, J., Sánchez, J. A., & Pérez, E. (2012). Acoustic markers associated with impairment in language processing in Alzheimer's disease. *The Spanish Journal of Psychology*, *15*(2), 487–494. https://doi.org/10.5209/rev_sjop.2012.v15.n2.38859

Migliore, S., Ghazaryan, A., Simonelli, I., Pasqualetti, P., Squitieri, F., Curcio, G., Landi, D., Palmieri, M. G., Moffa, F., Filippi, M. M., & Vernieri, F. (2017). Cognitive impairment in relapsing-remitting multiple sclerosis. Patients with very mild clinical disability. *Behavioural Neurology*, *2017*, 1–10. Article ID 7404289. https://doi.org/10.1155/2017/7404289

Moro-Velazquez, L., Cho, J. J., Watanabe, S., Hasegawa-Johnson, M. A., Scharenborg, O., Kim, H., & Dehak, N. (2019). Study of the performance of automatic speech recognition systems in speakers with Parkinson's disease. *Interspeech* (pp. 3875–3879). https://doi.org/10.21437/Interspeech.2019

Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M., & Beke, A. (2014). Development of a Large Spontaneous Speech Database of Agglutinative Hungarian Language. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds..), *Text, Speech and Dialogue. TSD 2014. Lecture Notes in Computer Science* (Vol. 8655, pp. 424–431). Springer.

Nocentini, U., Pasqualetti, P., Bonavita, S., Buccafusca, M., De Caro, M. F., Farina, D., Girlanda, P., Pira, F. L., Lugaresi, A., Quattrone, A., Reggio, A., Salemi, G., Savettieri, G., Tedeschi, G., Trojano, M., Valentino, P., & Caltagirone, C. (2006). Cognitive dysfunction in patients with relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, *12*(1), 77–87. https://doi.org/10.1191/135248506ms1227oa

Olivares, T., Nieto, A., Sánchez, M. P., Wollmann, T., Hernández, M. A., & Barroso, J. (2005). Pattern of neuropsychological impairment in the early phase of relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, *11*(2), 191–197. https://doi.org/10.1191/1352458505ms1139oa

Prakash, R. S., Snook, E. M., Lewis, J. M., Motl, R. W., & Kramer, A. F. (2008). Cognitive impairments in relapsing-remitting multiple sclerosis: A meta-analysis. *Multiple Sclerosis*, *14*(9), 1250–1261. https://doi.org/10.1177/1352458508095004

Prinster, A., Quarantelli, M., Orefice, G., Lanzillo, R., Brunetti, A., Mollica, C., Salvatore, E., Morra, V. B., Coppola, G., Vacca, G., Alfano, B., & Salvatore, M. (2006). Grey matter loss in relapsing-remitting multiple sclerosis: A voxel-based morphometry study. *NeuroImage*, *29*(3), 859–867. https://doi.org/10.1016/j.neuroimage.2005.08.034

Racsmány, M., Lukács, Á., Németh, D., & Pléh, C. (2005). A verbális munkamemória magyar nyelvű vizsgálóeljárásai. *Magyar Pszichológiai Szemle*, *LX*(4), 479–505. https://doi.org/10.1556/mpszle.60.2005.4.3

Rao, S. M. (2004). Cognitive function in patients with multiple sclerosis: Impairment and treatment. *International Journal of MS Care*, *6*(1), 9–22. https://doi.org/10.7224/1537-2073-6.1.9

Rocca, M. A., Valsasina, P., Martinelli, V., Misci, P., Falini, A., Comi, G., & Filippi, M. (2012). Large-scale neuronal network dysfunction in relapsing-remitting multiple sclerosis. *Neurology*, *79*(14), 1449–1457. https://doi.org/10.1212/WNL.0b013e31826d5f10

Rodgers, J. D., Tjaden, K., Feenaughty, L., Weinstock-Guttman, B., & Benedict, R. H. B. (2013). Influence of cognitive function on speech and articulation rate in multiple sclerosis. *Journal of International Neuropsychological Society*, *19*(2), 173–180. https://doi.org/10.1017/S1355617712001166

Sacco, R., Bisecco, A., Corbo, D., Della Corte, M., D'Ambrosio, A., Docimo, R., Gallo, A., Esposito, F., Esposito, S., Cirillo, M., Lavorgna, L., Tedeschi, G., & Bonavita, S. (2015). Cognitive impairment and memory disorders in relapsing–remitting multiple sclerosis: The role of white matter, gray matter and hippocampus. *Journal of Neurology*, *262*(7), 1691–1697. https://doi.org/10.1007/s00415-015-7763-y

Santiago, O., Guàrdia, J., Casado, V., Carmona, O., & Arbizu, T. (2007). Specificity of frontal dysfunctions in relapsing–remitting multiple sclerosis. *Archives of Clinical Neuropsychology*, *22*(5), 623–629. https://doi.org/10.1016/j.acn.2007.04.003

Satt, A., Hoory, R., König, A., Aalten, P., & Robert, P. H. (2014, October). Speech-based automatic and robust detection of very early dementia. *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech*, (pp. 2538–2542). https://doi.org/10.21437/interspeech.2014-544

Sidtis, D., Rogers, T., Godier, V., Tagliati, M., & Sidtis, J. J. (2015). Voice and fluency changes as a function of speech task and deep brain stimulation. *Journal of Language and Hearing Research*, *53*(5), 1167–1177. https://doi.org/10.1044/1092-4388(2010/09-0154)

Singman, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021) Afex: Analysis of factorial experiments. R package version 1.0-1. https://cran.r-project.org/package=afex

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and default mode: A quantitative meta-analysis. *Journal of cognitive neuroscience*, *21*(3), 489–510. https://doi.org/10.1162/jocn.2008.21029

Stachó, L., Dudás, R., Ivády, R., Kothencz, G., & Janka, Z. (2003). Addenbrooke's Kognitív vizsgálat: A magyar változat kifejlesztése. *Psychiatria Hungarica*, *18*(4), 226–240.

Steenwijk, M. D., Geurts, J. J. G., Daams, M., Tijms, B. M., Wink, A. M., Balk, L. J., Tewarie, P. K., Uitdehaag, B. M. J., Barkhof, F., Vrenken, H., & Pouwels, P. J. W. (2016). Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain*, *139*(1), 115–126. https://doi.org/10.1093/brain/awv337

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sumowski, J. F., Benedict, R., Enzinger, C., Filippi, M., Geurts, J. J., Hamalainen, P., Hulst, H., Inglese, M., Leavitt, V. M., Rocca, M. A., Rosti-Otajarvi, E. M., & Rao, S. (2018). Cognition in multiple sclerosis: State of the field and priorities for the future. *Neurology*, *90*(6), 278–288. https://doi.org/10.1212/WNL.0000000000004977

Svindt, V., Bóna, J., & Hoffmann, I. (2020). Changes in temporal features of speech in secondary progressive multiple sclerosis (Spms)–case studies. *Clinical Linguistics & Phonetics*, *34*(4), 339–356. https://doi.org/10.1080/02699206.2019.1645885

Tóth, L., Gosztolya, G., Vincze, V., Hoffmann, I., Szatlóczki, G., Biró, E., Zsura, F., Pákáski, M., & Kálmán, J. (2015). Automatic detection of mild cognitive impairment from spontaneous speech using ASR.

Tóth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatlóczki, G., Bánréti, Z., Pákáski, M., & Kálmán, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research*, *15*(2), 130–138. https://doi.org/10.2174/1567205014666171121114930

Van Lancker Sidtis, D., Cameron, K., & Sidtis, J. J. (2012). Dramatic effects of speech task on motor and linguistic planning in severely dysfluent parkinsonian speech. *Clinical Linguistics & Phonetics*, *26*(8), 695–711. https://doi.org/10.3109/02699206.2012.696307

Woods, S. P., Scott, J. C., Sires, D. A., Grant, I., Heaton, R. K., Tröster, A. I., Atkinson, J. H., McCutchan, J. A., Marcotte, T. D., Wallace, M. R., Ellis, R. J., Letendre, S., Schrier, R., Cherner, M., Sadek, J., Young, C., Jernigan, T., Lazzaretto, D. (2005). Action (verb) fluency: Test-retest reliability, normative standards, and construct validity. *Journal of the International Neuropsychological Society*, *11*(4), 408–415. https://doi.org/10.1017/S1355617705050460

Young, S., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Vatchev, V., & Woodland, P. (2006). *The HTK book*. Cambridge University Engineering Department.