Send Orders for Reprints to reprints@benthamscience.net

Current Alzheimer Research, XXXX, *XX*, 1-14 1

RESEARCH ARTICLE

Temporal Speech Parameters Detect Mild Cognitive Impairment in Different Languages: Validation and Comparison of the Speech-GAP Test® in English and Hungarian

János Kálmán[1,*], Davangere P. Devanand[2,3], Gábor Gosztolya[4], Réka Balogh[1], Nóra Imre[1], László Tóth[5], Ildikó Hoffmann[6,7], Ildikó Kovács[1], Veronika Vincze[4] and Magdolna Pákáski[1]

[1]Department of Psychiatry, Albert Szent-Györgyi Medical School, University of Szeged, Szeged, Hungary; [2]Department of Psychiatry, Columbia University Medical Center, New York, NY, United States; [3]Division of Geriatric Psychiatry, New York State Psychiatric Institute, New York, NY, United States; [4]MTA-SZTE Research Group on Artificial Intelligence, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary; [5]Institute of Informatics, Faculty of Science and Informatics, University of Szeged, Szeged, Hungary; [6]Department of Hungarian Linguistics, Faculty of Humanities and Social Sciences, University of Szeged, Szeged, Hungary; [7]Hungarian Research Centre for Linguistics, Eötvös Loránd Research Network, Budapest, Hungary

ARTICLE HISTORY

Received: December 20, 2021
Revised: February 08, 2022
Accepted: February 17, 2022

DOI:
10.2174/1567205019666220418155130

Abstract: ***Background***: The development of automatic speech recognition (ASR) technology allows the analysis of temporal (time-based) speech parameters characteristic of mild cognitive impairment (MCI). However, no information has been available on whether the analysis of spontaneous speech can be used with the same efficiency in different language environments.

Objective: The main goal of this international pilot study is to address the question of whether the Speech-Gap Test® (S-GAP Test®), previously tested in the Hungarian language, is appropriate for and applicable to the recognition of MCI in other languages such as English.

Methods: After an initial screening of 88 individuals, English-speaking (n = 33) and Hungarian-speaking (n = 33) participants were classified as having MCI or as healthy controls (HC) based on Petersen's criteria. The speech of each participant was recorded *via* a spontaneous speech task. Fifteen temporal parameters were determined and calculated through ASR.

Results: Seven temporal parameters in the English-speaking sample and 5 in the Hungarian-speaking sample showed significant differences between the MCI and the HC groups. Receiver operating characteristics (ROC) analysis clearly distinguished the English-speaking MCI cases from the HC group based on speech tempo and articulation tempo with 100% sensitivity, and on three more temporal parameters with high sensitivity (85.7%). In the Hungarian-speaking sample, the ROC analysis showed similar sensitivity rates (92.3%).

Conclusion: The results of this study in different native-speaking populations suggest that changes in acoustic parameters detected by the S-GAP Test® might be present across different languages.

Keywords: Mild cognitive impairment, neurocognitive disorder, language, speech analysis, temporal parameters, early recognition.

1. INTRODUCTION

Language changes occur in various types of neurocognitive disorders and are sensitive indicators of cortical dysfunction [1, 2]. The characteristic disruption in the language domain has been identified not only in different stages of dementia [3, 4], but also in its prodromal stage, mild cognitive impairment (MCI) [5]. However, recognition of the first clinical manifestations is still challenging since patients often do not recognize or minimize their deficits. In the early diagnostic procedure, there is an increasing need for non-invasive and cost-effective tools to identify individuals with minor neurocognitive disorders [4]. Since subtle changes in language and communication abilities may be apparent in the early course of such disorders [6], the detection of linguistic impairment could be a viable screening option [7-9]. Recordability of spoken language gives an opportunity to easily collect speech recordings, as biological samples. The purpose of our research group was to develop a new mobile application that would be capable of recording the examined person's telephone conversation and then analyzing the

*Address correspondence to this author at the Department of Psychiatry, Albert Szent-Györgyi Medical School, University of Szeged, Szeged, Hungary; Korányi fasor 8-10., H-6720, Szeged, Hungary; Tel: +36 62341600; E-mail: kalman.janos@med.u-szeged.hu

1567-2050/XX $65.00+.00

acoustic properties of spontaneous speech. Using this information technology (IT) technique, an individual can be examined through everyday activity, namely, a telephone conversation which is an ecologically valid way of assessment, decreases the time spent on neuropsychological tests, and eliminates test-induced anxiety for the user.

The first interest of our research team (beginning the now 10-year long research project on exploring the association between language function and cognition) was to identify speech parameters that might distinguish Hungarian patients with mild Alzheimer's disease (AD) from healthy controls (HC). Significant differences between the mild AD and the HC groups regarding speech tempo and hesitation ratio were first published by our research team [10]. However, in this early study, the transcription and annotation of speech signals were performed manually using the Praat software tool [11]. As the manual calculation of acoustic biomarkers is extremely time-consuming, its applicability in recognizing mild stages of cognitive deficits in clinical routine is rather limited.

However, the deterioration of acoustic language parameters can also be examined by implementing automatic speech recognition (ASR) techniques. ASR is a relatively simple and reliable method that has the potential to analyze large language datasets rapidly using machine-learning methods. Based on this technology, our research team developed the Speech-Gap Test®(S-GAP Test®) which identifies temporal (time-based) speech parameters using the extracted phonetic-level segmentation produced by ASR. In earlier studies applying the S-GAP Test®, we were able to distinguish MCI patients from HC subjects [12-17] based on several temporal parameters which demonstrated that the proposed acoustic characteristics indeed carry clinically relevant information in spontaneous speech [13].

Among the most informative temporal parameters, articulation and speech tempo, number and length of silent/filled pauses, and length of utterance were measured. Articulation/speech tempo is the number of phonemes per second during speech excluding/including hesitations, respectively. Hesitation is defined as the absence of speech and has two categories: silent pauses (silences that are not attributable to articulation constraints) and filled pauses (vocalizations like 'uhm', 'er'). A novelty of our studies was the focus on both silent and filled pauses along with the measurement of separated articulation and speech tempo. As our database of MCI patients was continuously growing and machine learning techniques were also exploited, the differentiation between MCI subjects and control probands gradually became more accurate (sensitivity: 81.3%; specificity: 66.7%) [16].

It is a basic requirement for diagnostic procedures used for the detection of MCI to be internationally applicable [18]. Particularly, in the case of procedures testing linguistic functions, the question arises of whether they have similar sensitivities in different languages. A recent systematic review emphasized that the methodology of speech-based studies in different native languages is quite heterogeneous [19]. Until now, phonetical-phonological analyses of speech for the assessment of cognitive impairment have been independently performed on native speakers of languages such as Chinese [20], English [21-27], French [28-31], Greek [32], Hungarian [13, 16, 17, 33, 34], Italian [35], Japanese [36-39], Persian [40], Spanish [2, 41-44], Swedish [45, 46], or Turkish [47]. However, until our present investigation, no information has been available on how the temporal characteristics of spontaneous speech compared between MCI *vs.* HC subjects in different language environments.

The main goal of this international pilot study was to explore the S-GAP-related temporal parameters of spontaneous speech in the English language with the purpose of MCI detection, and to address the methodological question of whether the S-GAP Test®, previously tested for Hungarian speakers, is appropriate for the recognition of speech parameters indicating MCI in the English language. Comparison of speech data obtained from native English- and Hungarian-speaking populations and assessing the effectiveness of the S-GAP Test® in these two different language environments would be the first step in the international application of this MCI-screening method. An IT application based on the S-GAP Test® could be a low-cost, non-invasive, and non-stressful method that could be applied in a rapid and easy way, without personal contact, and in a large population. The need for noncontact, remote assessment has also gained special urgency in light of the current COVID-19 pandemic.

2. MATERIALS AND METHODS

2.1. Participants and Study Design

Elderly individuals were recruited in parallel at two institutions: 1) Memory Disorders Center of the Department of Psychiatry, New York State Psychiatric Institute and Columbia University (New York, NY, USA) and 2) Memory Clinic, Department of Psychiatry, University of Szeged (Szeged, Hungary).

The ethnicity of the participants was not an inclusion or exclusion criterion and differed across the two study sites. The Hungarian participants were all Caucasian, while in the English-speaking group at Columbia University the individuals were Caucasian (69.7%), African-American (24.2%), and Hispanic (6.1%).

From the two outpatient clinics, a total of 88 individuals were recruited, 66 of whom were eligible for final inclusion (Fig. **1**). Both the English-speaking (n = 33) and Hungarian-speaking (n = 33) participants were classified as either MCI or as HC. The classification was based on Petersen's criteria [48] in both languages, with the Mini-Mental State Examination (MMSE) [49] serving as a measure for objective cognitive impairment (30-28 points: HC; 27-24 points: MCI).

To get an overview of participant characteristics and eligibility data, an interview focused on demographic features and medical history, as well as a brief neuropsychological test battery was administered (including the MMSE, the Clock Drawing Test (CDT) [50] and the Geriatric Depression Scale (GDS). All individuals were screened for possible dementia using the MMSE and those with a score under 24 were not involved in further participation. Corresponding to institutional protocols, the possibility of depression was also evaluated based on the 30-items [51] or the 15-items [52] version of the GDS (GDS-30/GDS-15; for the English-speaking/Hungarian-speaking sample, respectively): patients

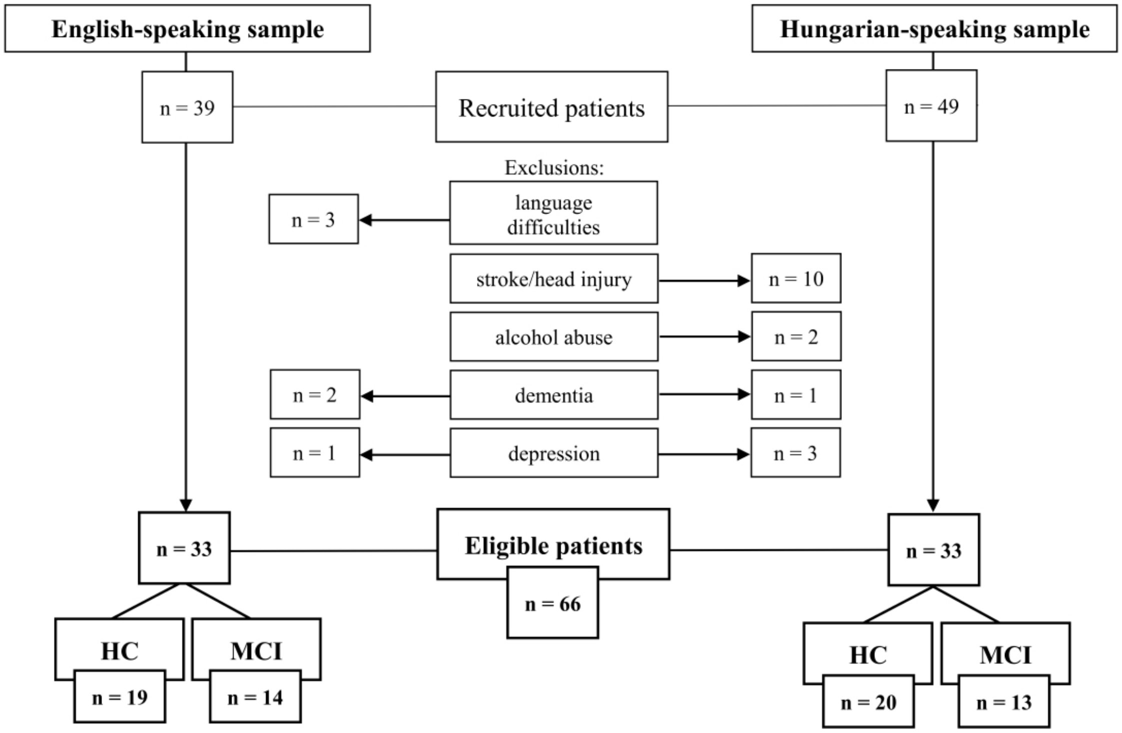

Fig. (1). Flowchart of participants' inclusion and exclusion process. **Abbreviations:** HC: healthy control; MCI: mild cognitive impairment.

scoring above 10 on GDS-30 or above 5 on GDS-15 were excluded.

Inclusion/exclusion criteria were the same at both sites. Inclusion criteria were a minimum age of 60 years, a minimum of 8 years of formal education, and English/Hungarian as native language (corresponding to the country of recruitment; bilingualism was not taken into account). Exclusion criteria included major hearing problems (*e.g.* uncorrected hearing loss), manifest speech problems (any form of aphasia), significant articulation problems (*e.g.* stutter), history of a substance use disorder, previous CT/MRI showing evidence of significant abnormality suggesting another potential etiology for MCI or dementia (*e.g.* micro- or macrohemorrhages, lacunar infarcts or single large infarct), evidence of cerebral contusion, encephalomalacia, aneurysm, vascular malformations, or clinically significant space-occupying lesions.

2.2. S-GAP Protocol and Preparation of Speech Samples

Following the clinical evaluation, speech samples were obtained from all participants. Spontaneous speech was elicited in the following way: Investigator 1, pointing to a mobile phone, informed the participant that a colleague (Investigator 2) would call from another room and provide instructions for a new task. Investigator 1 also told the participant that the conversation would be recorded and the task would only take a few minutes. Investigator 2 called the mobile phone, and after introduction, asked the participant to talk about his/her previous day. The standardized instruction was: *'Please tell me about your previous day in as much detail as you can.'* After the instruction, the investigators could not give verbal prompts, nor could they repeat the instruction; they remained silent throughout the call until the participant finished the task.

Each participants monologue was recorded by a call recorder application installed on the mobile phone device. The obtained recordings were then converted into an uncompressed PCM mono, 16-bit wav format with a sampling rate of 8,000 Hz. A professional expert linguist (I.H.) checked the quality of the recordings.

2.3. Analysis of Speech Samples

Pauses were defined as the disruption of speech for more than 30 ms (either silent segments in the case of silent-, or vocalizations in the case of filled pauses). Both silent and filled pauses were identified in each recording using ASR technology. Our ASR system was built on a modified version of the HTK tool [53], where we used the Hidden Markov model, but replaced the acoustic model with a Deep Neural Network (DNN) based one. This way, we utilized a standard HMM/DNN hybrid model, which is known to outperform traditional HMM models [54]. To realize the DNN acoustic model, we employed a custom DNN implementation [55] written in Visual C++ and utilized the CUDA library to speed up both model training and evaluation.

As acoustic features, we were using 40 raw Mel-frequency filter bank energy values along with the global log-energy, which was extended with the first and second-order derivatives ("FBANK + Δ + $\Delta\Delta$"), resulting in 123 acoustic features overall. Training and evaluation were done on a 150 ms (15 frames) wide sliding window, leading to 1,845 input neurons in the actual acoustic models. Then the acoustic model DNNs contained 5 fully connected hidden layers, each consisting of 1,024 neurons employing the

ReLU activation function [56], while they have a softmax final layer with a number of neurons equal to the phonetic units in the given language. The DNN acoustic models were trained for phoneme identification on two audio datasets consisting of spontaneous speech (as this type is expected to contain filled pauses), to match the language used by the subjects. For the speech samples in English, a subset of the TEDlium speech corpus [57] was used (100 speakers, approximately 15 hours of recording). For the Hungarian speech samples, a subset of the BEA corpus was employed (116 speakers, approximately 44 hours of recording) [58]. Before training, both corpora were down sampled to 8,000 Hz to match the sampling rate of the recordings in the study.

This ASR model was used to perform phoneme-level recognition, in which we also treated filled pauses as a special "phoneme". As language models, we employed simple phone bigrams both for English and Hungarian. This procedure produces a time-aligned phoneme sequence for each recording; that is, it supplies a hypothesis of the sequence of phones uttered, along with the starting and ending time indices. From this output, the 15 S-GAP parameters can be obtained *via* simple calculations (*e.g.* by counting the number of pauses and the total number of phones, and dividing the two values by each other; or by doing the same with the total duration of the pauses and all phones). We measured the accuracy of this workflow on a holdout set of the BEA corpus, consisting of 3 hours and 23 minutes, containing the speech of 10 subjects. Based on this, Pearson's correlation values of the speech tempo attributes calculated by our workflow and those derived from the transcripts were 0.857, while for articulation tempo this value was 0.866, indicating a precise (although not perfect) estimation. Most of the mismatching values were present in short speech segments: evaluating these values only for the segments with at least 2 seconds of duration led to Pearson's correlation values of 0.914 and 0.920, for articulation tempo and speech tempo, respectively. Furthermore, silent pauses were almost perfectly detected (precision: 96.1%, recall: 94.9%, F-measure: 95.5), while filled pauses were also identified with a high performance (precision: 83.2%, recall: 69.6%, F-measure: 75.8). In most cases, filled pauses were confused with prolongations of certain phonemes (*e.g.* m / n / a), which are acoustically similar and are often used by the speakers for similar purposes as filled pauses [59, 60].

The output of the ASR system was the phonetic segmentation and labeling of the input signal, which included filled pauses. Based on this output, we extracted 15 S-GAP-related temporal speech parameters using simple calculations (Table **1**).

2.4. Statistical Analysis

Descriptive statistics were used to examine the demographic, neuropsychological, and speech characteristics of participants. In both the English- and Hungarian-speaking samples, comparisons between the MCI *vs.* HC groups were executed using either the independent samples *t*-test/Welch's *t*-test (based on equality of variances), the Mann-Whitney U test (for cases when the normality assumption was not fulfilled according to the Shapiro-Wilk test of normality) or the Chi-square test (for categorical variables). For the examination of inter-language differences (English-speaking HC *vs.* Hungarian-speaking HC; English-speaking MCI *vs.* Hungarian-speaking MCI), independent samples *t*-test/Welch's *t*-test or the Mann-Whitney U test was carried out.

Table 1. List and definitions of the 15 S-GAP-related temporal parameters of spontaneous speech.

S-GAP-related Parameters	Description
Utterance length (s)	Total length of the utterance (s)
Articulation tempo (1/s)	Total number of phonemes (without hesitations) (count) / total length of the utterance (s)
Speech tempo (1/s)	Total number of phonemes (including hesitations) (count) / total length of the utterance (s)
Silent pause occurrence rate (%)	Total number of silent pauses (count) x 100 / total number of phonemes (count)
Filled pause occurrence rate (%)	Total number of filled pauses (count) x 100 / total number of phonemes (count)
Total pause occurrence rate (%)	Total number of silent and filled pauses (count) x 100 / total number of phonemes (count)
Silent pause duration rate (%)	Total length of silent pauses (s) x 100 / total length of the utterance (s)
Filled pause duration rate (%)	Total length of filled pauses (s) x 100 / total length of the utterance (s)
Total pause duration rate (%)	Total length of silent and filled pauses (s) x 100 / total length of the utterance (s)
Silent pause frequency (1/s)	Total number of silent pauses (count) / total length of the utterance (s)
Filled pause frequency (1/s)	Total number of filled pauses (count) / total length of the utterance (s)
Total pause frequency (1/s)	Total number of silent and filled pauses (count) / total length of the utterance (s)
Silent pause average duration (s)	Total length of silent pauses (s) / total number of silent pauses (count)
Filled pause average duration (s)	Total length of filled pauses (s) / total number of filled pauses (count)
Total pause average duration (s)	Total length of silent and filled pauses (s) / total number of silent and filled pauses (count)

Abbreviations: s: second.

Receiver operating characteristics (ROC) analysis was applied to assess which S-GAP-related parameters have the most promising classification abilities based on their area under the curve (AUC) in the two languages. Sensitivity and specificity (true positive rate and true negative rate) were calculated using threshold values that yielded the highest possible sensitivity (while keeping specificity above 50%). For comparison of the S-GAP parameters' classification ability between the two languages, the comparison of the independent ROC curves module of the MedCalc software was used.

All statistical analyses were performed using SPSS v.24 (SPSS Inc., Chicago, IL, USA), except for the inter-language comparison of AUCs for which MedCalc v.19.4 was applied (MedCalc Software Ltd., Ostend, Belgium). For all statistical comparisons, the level of significance was set at the 0.05 level.

3. RESULTS

3.1. Demographics and Neuropsychological Test Performances

Detailed demographic characteristics and neuropsychological test scores of all groups (means and standard deviations) are presented in Table **2**. Regarding demographics (gender, age, and years of education) and the CDT test, there were no statistically significant differences between the MCI and the HC group in either languages. However, regarding the other neuropsychological tests, MCI patients showed significantly poorer performance in the MMSE than HCs (English-speaking sample: U = 62.500; Z = -2.703; p = 0.009; Hungarian-speaking sample: U = 0.000; Z = -4.879; p < 0.001), and they also had higher scores in the GDS in both languages (English-speaking sample: U = 71.000; Z = -2.277; p = 0.024; Hungarian-speaking sample: U = 59.000; Z = -2.736; p = 0.008).

3.2. S-GAP-related Temporal Parameters and Sensitivity Measures in the English-Speaking Sample

Regarding the English-speaking sample, 7 of the total 15 S-GAP-related temporal parameters displayed significant differences between the MCI and the HC groups. Patients with MCI showed significantly lower articulation tempo and speech tempo as well, while they produced a significantly higher occurrence rate of total pauses, duration rate of silent pauses and total pauses, as well as the average duration of silent pauses and total pauses (Table **3**).

To determine which S-GAP-related temporal speech parameters would be the most precise in classifying patients, ROC analysis was executed. The ROC analysis revealed that the following 8 parameters had statistically significant classification abilities (starting with the highest AUC): speech tempo, articulation tempo, total pause duration rate, silent pause duration rate, silent pause average duration, total pause average duration, total pause occurrence rate, and filled pause occurrence rate. Sensitivity was above 90% both for speech tempo (sensitivity: 100%; specificity: 63.2%) and for articulation tempo (sensitivity: 100%; specificity: 57.9%).

Sensitivity and specificity measures of the statistically significant S-GAP-related temporal parameters (calculated using threshold values optimal for early screening) are detailed in Table **4**; ROC curves are plotted in Fig. (**2**).

3.3. S-GAP-related Temporal Parameters and Sensitivity Measures in the Hungarian-Speaking Sample

Regarding the Hungarian-speaking sample, 5 of the total 15 S-GAP-related temporal parameters turned out to be statistically different between the MCI and the HC group. MCI patients' utterance length was significantly shorter, while a higher duration rate of silent pauses and total pauses, as well as a higher average duration of silent pauses and total pauses characterized their speech (Table **5**).

With regard to the ROC analysis, the following 5 parameters turned out to be statistically significant (from highest to lowest AUCs): silent pause duration rate, utterance length, total pause duration rate, silent pause average duration, and total pause average duration. Sensitivity was above 90% in

Table 2. Means (standard deviations) of participants' demographic characteristics and neuropsychological test scores in the English-speaking and Hungarian-speaking samples.

English-Speaking Sample		-	Hungarian-Speaking Sample	
HC (**n = 19**)	**MCI** (**n = 14**)	-	**HC** (**n = 20**)	**MCI** (**n = 13**)
-		**Demographic characteristics**	-	
5/14	6/8	**Gender (male/female)**	3/17	4/9
74.47 (7.321)	72.36 (6.857)	**Age (years)**	69.90 (5.609)	73.77 (4.969)
17.84 (3.532)	16.79 (3.118)	**Education (years)**	13.15 (2.455)	11.77 (2.743)
-		**Neuropsychological test scores**	-	
29.16 (1.015)	27.71 (1.773)	**MMSE**	28.85 (0.813)	26.31 (0.751)
8.89 (1.197)	9.21 (1.188)	**CDT**	7.60 (3.152)	7.92 (2.178)
3.16 (2.853)	5.50 (2.822)	**GDS-30 / GDS-15**	1.65 (1.387)	2.77 (1.013)

Abbreviations: HC: healthy control; MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; CDT: Clock Drawing Test; GDS-30: Geriatric Depression Scale (30-item); GDS-15: Geriatric Depression Scale (15-item).

Table 3. Descriptive statistics (means and standard deviations) and group comparisons in the English-speaking sample using the independent samples *t*-test / Mann-Whitney U test.

English-Speaking Sample	M (SD)		Test STATISTICS	
S-GAP-Related Parameters	**HC (n = 19)**	**MCI (n = 14)**	***t*-test / Mann-Whitney U TEST**	***p***
Utterance length (s)	275.33 (120.02)	201.94 (135.07)	U = 82.000; Z = -1.858	0.065
Articulation tempo (1/s)	8.88 (1.21)	6.78 (1.32)	$t(31)$ = 4.732	0.000*
Speech tempo (1/s)	10.07 (1.10)	8.02 (1.34)	$t(31)$ = 4.810	0.000*
Silent pause occurrence rate (%)	9.43 (3.17)	12.11 (4.35)	U = 85.000; Z = -1.748	0.084
Filled pause occurrence rate (%)	2.55 (1.08)	3.63 (1.73)	U = 79.000; Z = -1.967	0.050
Total pause occurrence rate (%)	11.98 (3.55)	15.75 (4.34)	$t(31)$ = -2.736	0.010*
Silent pause duration rate (%)	31.43 (8.72)	45.61 (12.05)	$t(31)$ = -3.927	0.000*
Filled pause duration rate (%)	5.64 (3.23)	6.56 (5.22)	U = 126.000; Z = -0.255	0.815
Total pause duration rate (%)	37.07 (9.27)	52.17 (11.23)	$t(31)$ = -4.228	0.000*
Silent pause frequency (1/s)	0.93 (0.30)	0.95 (0.28)	$t(31)$ = -0.139	0.890
Filled pause frequency (1/s)	0.25 (0.09)	0.28 (0.14)	U = 122.000; Z = -0.401	0.706
Total pause frequency (1/s)	1.18 (0.33)	1.24 (0.30)	$t(31)$ = -0.453	0.653
Silent pause average duration (s)	0.34 (0.07)	0.51 (0.18)	$t(15.802)$ = -3.108	0.007*
Filled pause average duration (s)	0.21 (0.05)	0.21 (0.09)	U = 105.000;Z = -1.020	0.321
Total pause average duration (s)	0.31 (0.05)	0.44 (0.14)	$t(15.968)$ = -3.007	0.008*

Abbreviations: M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment; *p-values indicating statistically significant differences (level of significance was set at $p < 0.05$).

Table 4. Accuracy measures of S-GAP-related temporal parameters with statistically significant classification ability in the English-speaking sample using ROC analysis.

English-Speaking Sample	Accuracy Measures						
S-GAP-Related Parameters	***p***	**AUC**	**95% CI-**	**95% CI+**	**Threshold Value**	**Sensitivity (%)**	**Specificity (%)**
Speech tempo (1/s)	0.000	0.891	0.784	0.998	9.843	100	63.2
Articulation tempo (1/s)	0.000	0.891	0.779	1.000	8.772	100	57.9
Total pause duration rate (%)	0.001	0.846	0.711	0.980	36.689	85.7	52.6
Silent pause duration rate (%)	0.001	0.835	0.695	0.974	32.398	85.7	63.2
Silent pause average duration (s)	0.003	0.808	0.654	0.963	0.346	85.7	52.6
Total pause average duration (s)	0.006	0.782	0.614	0.950	0.329	78.6	57.9
Total pause occurrence rate (%)	0.016	0.748	0.578	0.918	12.078	78.6	52.6
Filled pause occurrence rate (%)	0.049	0.703	0.524	0.882	2.567	78.6	52.6

Abbreviations: ROC: receiver operating characteristics; AUC: area under the curve; CI: confidence interval; (level of significance was set at $p < 0.05$).

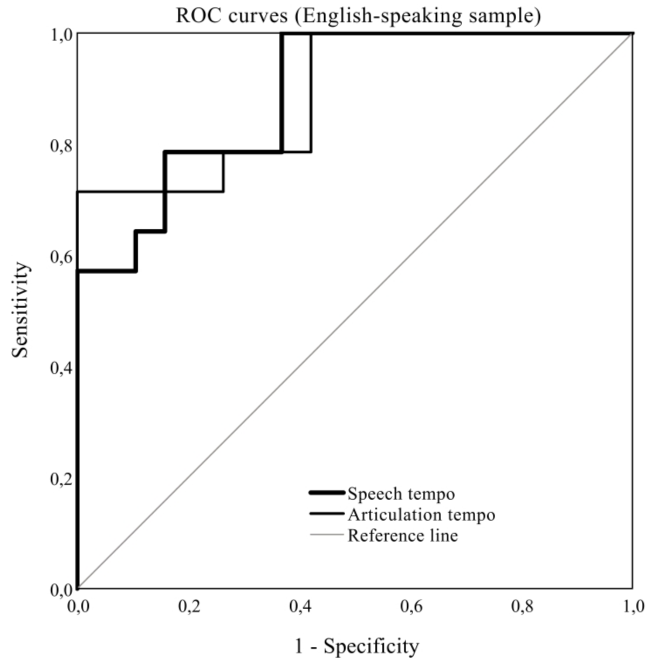

Fig. (2). ROC curves for S-GAP-related temporal parameters with the highest (above 90%) sensitivity for discriminating between MCI and HC participants in the English-speaking sample (speech tempo and articulation tempo). **Abbreviations:** ROC: receiver operating characteristics; HC: healthy control; MCI: mild cognitive impairment.

Table 5. Descriptive statistics (means and standard deviations) and group comparisons in the Hungarian-speaking sample using the independent samples *t*-test / Mann-Whitney U test.

Hungarian-Speaking Sample	**M (SD)**		**Test Statistics**	
S-GAP-Related Parameters	**HC (*n* = 20)**	**MCI (*n* = 13)**	***t*-test / Mann-Whitney U Test**	***p***
Utterance length (s)	155.06 (70.21)	107.82 (87.65)	U = 66.000; Z = -2.358	0.018*
Articulation tempo (1/s)	9.90 (1.97)	8.63 (1.75)	$t(31)$ = 1.878	0.070
Speech tempo (1/s)	10.67 (1.87)	9.47 (1.62)	$t(31)$ = 1.894	0.068
Silent pause occurrence rate (%)	4.88 (1.64)	5.91 (1.83)	$t(31)$ = -1.678	0.103
Filled pause occurrence rate (%)	2.69 (1.83)	3.28 (2.10)	U = 112.000; Z = -0.663	0.524
Total pause occurrence rate (%)	7.58 (3.13)	9.20 (3.37)	U = 94.500; Z = -1.308	0.194
Silent pause duration rate (%)	23.49 (9.72)	32.46 (8.16)	$t(31)$ = -2.750	0.010*
Filled pause duration rate (%)	6.26 (4.10)	7.03 (4.68)	$t(31)$ = -0.494	0.625
Total pause duration rate (%)	29.76 (11.81)	39.49 (11.07)	$t(31)$ = -2.367	0.024*
Silent pause frequency (1/s)	0.49 (0.11)	0.54 (0.13)	$t(31)$ = -1.008	0.321
Filled pause frequency (1/s)	0.26 (0.14)	0.28 (0.15)	$t(31)$ = -0.336	0.739
Total pause frequency (1/s)	0.76 (0.21)	0.83 (0.22)	U = 108.000; Z = -0.811	0.434
Silent pause average duration (s)	0.47 (0.18)	0.62 (0.17)	U = 70.000; Z = -2.211	0.027*
Filled pause average duration (s)	0.21 (0.06)	0.24 (0.10)	U = 123.000; Z = -0.258	0.813
Total pause average duration (s)	0.39 (0.14)	0.48 (0.10)	U = 73.000; Z = -2.100	0.036*

Abbreviations: M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment; *p-values indicating statistically significant differences (level of significance was set at $p < 0.05$).

Table 6. Accuracy measures of S-GAP-related temporal parameters with statistically significant classification ability in the Hungarian-speaking sample using ROC analysis.

Hungarian-Speaking Sample	Accuracy Measures						
S-GAP-Related Parameters	p	AUC	95% CI-	95% CI+	Threshold Value	Sensitivity (%)	Specificity (%)
Silent pause duration rate (%)	0.018	0.746	0.579	0.914	24.191	92.3	60.0
Utterance length (s)	0.018	0.746	0.558	0.934	132.345	76.9	60.0
Total pause duration rate (%)	0.020	0.742	0.573	0.912	27.280	92.3	55.0
Silent pause average duration (s)	0.027	0.731	0.551	0.910	0.438	84.6	55.0
Total pause average duration (s)	0.036	0.719	0.537	0.902	0.349	92.3	55.0

Abbreviations: ROC: receiver operating characteristics; AUC: area under the curve; CI: confidence interval; (level of significance was set at $p < 0.05$).

the case of three parameters, with the highest specificity for silent pause duration rate (sensitivity: 92.3%; specificity: 60.0%) while lower for total pause duration rate (sensitivity: 92.3%; specificity: 55.0%), and total pause average duration (sensitivity: 92.3%; specificity: 55.0%).

Sensitivity and specificity measures of the statistically significant temporal parameters (calculated at optimal threshold values) are detailed in Table **6**; ROC curves are plotted in Fig. (**3**).

To examine whether the S-GAP-related parameters have different classification abilities in the two languages, pair-wise comparisons of AUCs were executed between the English- and Hungarian-speaking samples. The analysis showed that the AUCs did not differ significantly regarding any of the 15 S-GAP-related parameters between the two language groups (Table **7**).

3.4. Inter-Language Group Comparisons of S-GAP-related Temporal Parameters

Besides our main goal of exploring the S-GAP-related temporal parameters separately in the two language samples, inter-language comparisons were also carried out as additional analyses between the English-speaking *vs.* Hungarian-speaking HC group and the English-speaking *vs.* Hungarian-speaking MCI group (Table **8**). Regarding the HC group, 8 S-GAP-related parameters showed statistically significant differences between the English- and Hungarian-speaking samples, which were the following: utterance length (E-HC > H-HC), silent pause occurrence rate (E-HC > H-HC), total pause occurrence rate (E-HC > H-HC), silent pause duration rate (E-HC > H-HC), total pause duration rate (E-HC > H-HC), silent pause frequency (E-HC > H-HC), total pause frequency (E-HC > H-HC), and silent pause average duration (H-HC > E-HC). Regarding the MCI group, 9 significantly different parameters were revealed again: utterance length (E-MCI > H-MCI), articulation tempo (H-MCI > E-MCI), speech tempo (H-MCI > E-MCI), silent pause occurrence rate (E-MCI > H-MCI), total pause occurrence rate (E-MCI > H-MCI), silent pause duration rate (E-MCI > H-MCI), total pause duration rate (E-MCI > H-MCI), silent pause frequency (E-MCI > H-MCI), and total pause frequency (E-MCI > H-MCI).

4. DISCUSSION

The aim of this international study was to validate the S-GAP Test®, a novel spontaneous speech analyzer (originally developed for the Hungarian language), in an English-speaking sample for the purpose of MCI-recognition. The major objective was to develop a neuropsychological screening method, which is sensitive to in multiple languages and provides clinicians with a simple and quick way for the screening of MCI. For this purpose, automatic analysis of spontaneous speech was carried out by applying ASR. This is the first study conducted with both English- and Hungarian native speakers in which the same method was applied to explore the acoustic parameters of spontaneous speech in MCI and HC subjects.

To summarize the 10-year development process of the S-GAP Test®, the first main finding was the discovery of significant differences between the mild stage of AD and HC regarding speech tempo and hesitation ratio [10]; subsequently, its usefulness was also demonstrated in the prodromal stage of AD since the proposed acoustic biomarkers carried significant information on the separation of MCI from HC [13]. In parallel with the introduction of MCI as a target group, another important step in the development process was the implementation of automatic analysis instead of manual counting. Through the efforts toward the automatic extraction of acoustic features, a machine learning model was constructed [13, 15]. The automatically selected feature sets were found to be superior to the manually constructed ones used for MCI detection [14]. Extending the previous studies, the applicability of the S-GAP Test® was demonstrated in differentiation not only between MCI and HC but also between MCI and mild AD patients by relying on automatically extracted acoustic markers of spontaneous speech [17, 33]. Before the present study, the S-GAP Test® was applied to a total of 95 HC, 105 MCI, and 35 mild AD individuals.

4.1. Main Findings

Present results indicated that analysis of spontaneous speech using the S-GAP Test® is sensitive to detect MCI cases not only in native Hungarian-speaking but in native English-speaking populations as well. Four temporal parameters that differed significantly between the HC and MCI groups both in the English-speaking and in the Hungarian-

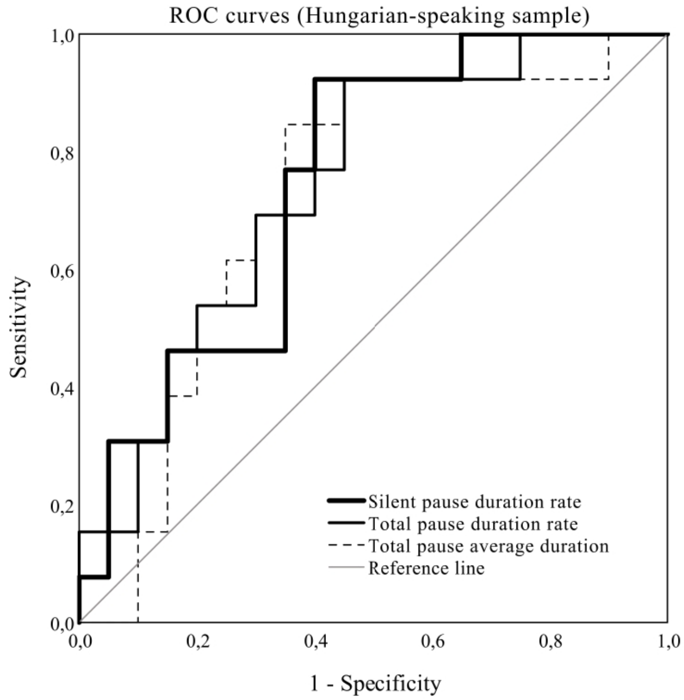

Fig. (3). ROC curves for temporal parameters with the highest (above 90%) sensitivity for discriminating between MCI and HC participants in the Hungarian-speaking sample (silent pause duration rate, total pause duration rate, and total pause average duration). **Abbreviations:** ROC: receiver operating characteristics; HC: healthy control; MCI: mild cognitive impairment.

Table 7. Pairwise comparison of the English- and Hungarian-speaking samples' AUCs regarding the 15 S-GAP-related temporal parameters of speech.

-	AUC		Pairwise Comparisons	
S-GAP-Related Parameters	**English-Speaking Sample**	**Hungarian-Speaking Sample**	***z- statistic***	***p***
Utterance length (s)	0.692	0.746	0.384	0.701
Articulation tempo (1/s)	0.891	0.692	1.741	0.082
Speech tempo (1/s)	0.891	0.685	1.828	0.068
Silent pause occurrence rate (%)	0.680	0.658	0.163	0.871
Filled pause occurrence rate (%)	0.703	0.569	0.931	0.352
Total pause occurrence rate (%)	0.748	0.637	0.827	0.408
Silent pause duration rate (%)	0.835	0.746	0.784	0.433
Filled pause duration rate (%)	0.528	0.508	0.120	0.904
Total pause duration rate (%)	0.846	0.743	0.927	0.354
Silent pause frequency (1/s)	0.541	0.631	0.600	0.548
Filled pause frequency (1/s)	0.541	0.523	0.119	0.905
Total pause frequency (1/s)	0.560	0.585	0.169	0.866
Silent pause average duration (s)	0.808	0.731	0.630	0.529
Filled pause average duration (s)	0.605	0.527	0.492	0.623
Total pause average duration (s)	0.782	0.719	0.486	0.627

Abbreviations: AUC: area under the curve; (level of significance was set at $p < 0.05$).

Table 8. Inter-language comparisons of the S-GAP-related temporal parameters of speech using the independent samples *t*-test / Mann-Whitney U test.

S-GAP-Related Parameters	E-HC *vs.* H-HC		E-MCI *vs.* H-MCI	
	t-test / Mann-Whitney U Test	*p*	*t*-test / Mann-Whitney U Test	*p*
Utterance length (s)	$t(28.729) = 3.794$	0.001*	$U = 47.000$; $Z = -2.135$	0.033*
Articulation tempo (1/s)	$t(31.801) = -1.949$	0.060	$t(25) = -3.120$	0.005*
Speech tempo (1/s)	$t(31.081) = -1.219$	0.232	$t(25) = -2.529$	0.018*
Silent pause occurrence rate (%)	$t(26.715) = 5.570$	0.000*	$U = 8.000$; $Z = -4.028$	0.000*
Filled pause occurrence rate (%)	$U = 179.000$; $Z = -0.309$	0.771	$U = 74.000$; $Z = -0.825$	0.430
Total pause occurrence rate (%)	$U = 65.000$; $Z = -3.512$	0.000*	$t(25) = 4.347$	0.000*
Silent pause duration rate (%)	$t(37) = 2.678$	0.011*	$t(25) = 3.293$	0.003*
Filled pause duration rate (%)	$U = 174.000$; $Z = -0.450$	0.667	$U = 85.000$; $Z = -0.291$	0.793
Total pause duration rate (%)	$t(37) = 2.142$	0.039*	$t(25) = 2.951$	0.007*
Silent pause frequency (1/s)	$t(23.309) = 5.898$	0.000*	$t(25) = 4.652$	0.000*
Filled pause frequency (1/s)	$t(37) = -0.400$	0.691	$U = 89.000$; $Z = -0.097$	0.943
Total pause frequency (1/s)	$U = 55.000$; $Z = -3.793$	0.000*	$U = 17.000$; $Z = -3.591$	0.000*
Silent pause average duration (s)	$U = 110.000$; $Z = -2.248$	0.024*	$t(25) = -1.537$	0.137
Filled pause average duration (s)	$U = 148.000$; $Z = -1.180$	0.247	$U = 73.000$; $Z = -0.873$	0.402
Total pause average duration (s)	$U = 141.000$; $Z = -1.377$	0.175	$t(25) = -0.789$	0.437

Abbreviations: E-HC: English-speaking sample - healthy control; E-MCI: English-speaking sample - mild cognitive impairment; H-HC: Hungarian-speaking sample - healthy control; H-MCI: Hungarian-speaking sample - mild cognitive impairment; **p*-values indicating statistically significant differences (level of significance was set at $p< 0.05$).

speaking samples are: MCI patients showed higher silent pause duration rate, total pause duration rate, silent pause average duration, and total pause average duration. Based on this finding, these parameters might be sensitive biomarkers of MCI in both languages.

Additional to the above-mentioned four temporal parameters, the English-speaking MCI group also showed lower articulation tempo and speech tempo compared to HC. The importance of these linguistic features in mild AD or MCI has been previously demonstrated in the Hungarian language, using both manual calculation and automatic analysis [10, 13, 16, 17, 33]. Interestingly, in our previous studies, Hungarian-speaking MCI/AD patients also showed a reduction in articulation and speech tempo, while in the present sample this difference was only tendentious. A possible explanation of this might be the difference in the task that was implemented for speech elicitation: namely, in our previous studies, a film description task was used in which the participants had to retell the events of a specially designed, one-minute long silhouette animation, instead of the previous day task applied in the present study.

ROC analysis clearly distinguished the English-speaking MCI cases from HCs based on speech tempo and articulation tempo with 100% sensitivity and further three parameters with very high sensitivity (85.7%) at moderate specificity. In the Hungarian-speaking groups, ROC analysis showed high sensitivity values for silent and total pause duration rate and also for total pause average duration (92.3%). These results suggest that the S-GAP Test® might indicate MCI more sensitively in the English-speaking than in the Hungarian-speaking sample.

Higher number and/or length of pauses, and the decrease of articulation/speech tempo have been described in a number of studies examining varying degrees of cognitive impairment, however, with different methodologies and using various types of tasks such as spontaneous speech [61-63], narrative recall [21], picture description [32], or reading aloud [2, 24, 43].

Pause-related features indicate retrieval difficulties [12] related to degeneration in hippocampal brain regions [64], and they are also associated with atrophy of grey matter in the frontopolar (or Brodmann) area [65] which has a role in higher-order cognitive functions like memory retrieval [66] or multitasking [67]. It is hypothesized that an increase in the number or duration of pauses demonstrates the increase in the cognitive load required for maintaining one's train of thought during speech [28]. Although these changes might not always be perceptible to the ear, speech analysis indicates that silence might be a significant marker of planning, word-retrieval, and executive difficulties due to cognitive deterioration [2, 61]. Language functions in general (*e.g.* measured by naming or verbal fluency tasks) also show a correlation with grey matter volume of the left temporal lobe in MCI and AD [68].

It is important to note that metrics regarding the diagnostic accuracy of language functions have been reported in variants of primary progressive aphasia [69] but earlier investigations of MCI/AD did not focus on this [21, 22]. However, in recent years more data related to this field have been reported [16, 17, 28, 29, 33, 70]. For example, the classification sensitivity of linguistic and phonetic features of connected speech by automated assessment of the Cookie Theft picture description task was 85% between HC and AD/MCI cases in a Canadian-Mexican co-operation study [70]. The diagnostic utility of automatic speech analysis for recorded vocal tasks has also been previously demonstrated in a French-speaking population with 79% or 86% classification accuracy between HC and MCI [28, 29]. As for the previous investigations of our research group, the irregularities in MCI speech and language were demonstrated by 68% sensitivity in the differentiating MCI from HC [17]. This result, however, was based on a different language elicitation task, *i.e.* a video description (as mentioned earlier). Comparing results of the present study with previous ones, the S-GAP Test® applied for English- and Hungarian-speaking MCI populations has shown a relatively high sensitivity.

4.2. Limitations and Considerations

A limitation of this pilot study was the small sample size and particularly the low number of MCI participants, which represent the main drawback regarding statistical power. However, this disadvantage was compensated by careful examination of the patients included in the study with the aim of excluding other confounding factors. Taking into consideration that this research was intended as a pilot to find those temporal parameters (from the full set of 15) with the highest differentiating potential for embedding in a future mobile application, multiple correction testing was not applied for the statistical comparisons. This needs to be taken into account when interpreting the results.

Regarding the sensitivity and specificity of temporal parameters, the optimal threshold values were defined to maximize sensitivity, which, as a result of a trade-off between the two measures, decreased specificity (although it exceeded 50% in every case). Given that the goal was to create an early MCI screening tool specifically targeting high-risk individuals (*e.g.* people above the age of 60) and considering the serious consequences of undiagnosed MCI (mainly the possibility of converting to dementia [71], reaching a high true positive rate was prioritized.

Before applying the S-GAP Test® internationally in clinical settings, the observed inter-language differences (E-HC *vs.* H-HC; E-MCI *vs.* H-MCI) emphasize the need for gathering normative data for international adaptations. In our present sample, English-speaking individuals on average produced longer monologues regarding their previous day, while they talked slower and their speech contained more pauses compared to the Hungarian-speaking participants. These language differences will have to be taken into account during the setting of screening thresholds in different countries as temporal features indicative of HC/MCI speech can have substantially different mean values in each language.

CONCLUSION

In summary, the results of the S-GAP Test® in the English- and Hungarian native speaker populations suggest that similar changes in temporal parameters of spontaneous speech detected by ASR can be observed across different languages. Based on these findings, it could be suggested that the S-GAP Test® has the potential to become a useful method for early MCI screening both in English-speaking and Hungarian-speaking populations. An early and accurate diagnosis of cognitive deficits would be of much help for patients and their families in order to plan for the future and to start early treatment. However, it is important to state that this method can only be the first step in the diagnostic process of MCI, as it is not intended to be a complete substitute for a detailed clinical examination.

In the future, an S-GAP Test®-based speech analysis might permit the screening and research evaluation of prodromal stages of different types of dementia through a computerized, interactive smart phone application (which is currently under development in co-operation with the Institute of Informatics at the University of Szeged, Hungary). This could be a low-cost, noninvasive, non-stressful method that allows quick, easy, and remote assessment. A further advantage of this method is that the recording of spontaneous speech (in a phone call-like setting) is less stressful for the patient than a neuropsychiatric test. Additionally, this approach might also serve as an objective measurement for the efficacy of pharmacotherapy and drug candidate molecules in cognitive impairment.

LIST OF ABBREVIATIONS

AD	=	Alzheimer's Disease
ASR	=	Automatic Speech Recognition
AUC	=	Area Under Curve
CDT	=	Clock Drawing Test
CT	=	Computed Tomography
DNN	=	Deep Neural Network
GDS	=	Geriatric Depression Scale
HC	=	Healthy Control
HMM	=	Hidden Markov Model
IT	=	Information Technology
M	=	Mean
MCI	=	Mild Cognitive Impairment
MRI	=	Magnetic Resonance Imaging
MMSE	=	Mini-Mental State Examination
ROC	=	Receiver Operating Characteristic
s	=	Second
SD	=	Standard Deviation
S-GAP Test®	=	Speech-Gap Test®

AUTHORS' CONTRIBUTION

J.K. and I.H. conceived and designed the study. D.P.D. and M.P. evaluated the subjects, while I.K., R.B. and N.I. collected the data. R.B., N.I., G.G., L.T., V.V. analyzed the data and performed the figures and tables. J.K., M.P., G.G., R.B. and N.I. wrote the manuscript. All authors contributed to the article and approved the final manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The American part of the study was approved by the Institutional Review Board of the New York State Psychiatric Institute – Columbia University Department of Psychiatry (protocol number: 7611). The Hungarian part of the study was approved by the Regional Human Biomedical Research Ethics Committee of the University of Szeged, Hungary (reference number: 231/2017-SZTE).

HUMAN AND ANIMAL RIGHTS

No animals were used in this study. All human research procedures followed were in accordance with the guidelines of the Declaration of Helsinki of 1975, as revised in 2013 (http://ethics.iit.edu/ecodes/node/3931).

CONSENT FOR PUBLICATION

Written informed consent was obtained from all participants at both research sites.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

This work was supported by Merck & Co., Inc., the National Research, Development, and Innovation Office of Hungary (NKFIH-FK-124413), and the Hungarian Ministry for Innovation and Technology (TUDFO/47138-1/2019-ITM). The authors were also supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (G. Gosztolya: BO/00653/19), and the Hungarian Ministry for Innovation and Technology (G. Gosztolya: NKFIH-1279-2/2020, NKFIH-FK-124413, ÚNKP-21-5; N. Imre: ÚNKP-21-4-I).

CONFLICT OF INTEREST

Dr. D. P. Devanand is a consultant to Acadia, BXCel, Genentech, Corium, and Grifols. The authors have no competing interests to declare.

ACKNOWLEDGEMENTS

The authors wish to thank Nóra Domján for her help in performing the neuropsychiatric tests and recording the audio files of the participants' speeches.

REFERENCES

[1] Braaten AJ, Parsons TD, McCue R, Sellers A, Burns WJ. Neurocognitive differential diagnosis of dementing diseases: Alzheimer's dementia, vascular dementia, frontotemporal dementia, and major depressive disorder. Int J Neurosci 2006; 116(11): 1271-93. http://dx.doi.org/10.1080/00207450600920928 PMID: 17000529

[2] Meilán JJG, Martínez-Sánchez F, Carro J, Sánchez JA, Pérez E. Acoustic markers associated with impairment in language processing in Alzheimer's Disease. Span J Psychol 2012; 15(2): 487-94. http://dx.doi.org/10.5209/rev_SJOP.2012.v15.n2.38859 PMID: 22774422

[3] Forbes KE, Venneri A, Shanks MF. Distinct patterns of spontaneous speech deterioration: An early predictor of Alzheimer's disease. Brain Cogn 2002; 48(2-3): 356-61. http://dx.doi.org/10.1006/brcg.2001.1377 PMID: 12030467

[4] Laske C, Sohrabi HR, Frost SM, *et al.* Innovative diagnostic tools for early detection of Alzheimer's disease. Alzheimers Dement 2015; 11(5): 561-78. http://dx.doi.org/10.1016/j.jalz.2014.06.004 PMID: 25443858

[5] Taler V, Phillips NA. Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. J Clin Exp Neuropsychol 2008; 30(5): 501-56. http://dx.doi.org/10.1080/13803390701550128 PMID: 18569251

[6] Ahmed S, Haigh AM, de Jager CA, Garrard P. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. Brain 2013; 136(Pt 12): 3727-37. http://dx.doi.org/10.1093/brain/awt269 PMID: 24142144

[7] Petti U, Baker S, Korhonen A. A systematic literature review of automatic Alzheimer's disease detection from speech and language. J Am Med Inform Assoc 2020; 27(11): 1784-97. http://dx.doi.org/10.1093/jamia/ocaa174 PMID: 32929494

[8] Martínez-Nicolás I, Llorente TE, Martínez-Sánchez F, Meilán JJG. Ten years of research on automatic voice and speech analysis of people with Alzheimer's disease and mild cognitive impairment: A systematic review article. Front Psychol 2021; 12: 620251. http://dx.doi.org/10.3389/fpsyg.2021.620251 PMID: 33833713

[9] Vigo I, Coelho L, Reis S. Speech- and language-based classification of Alzheimer's disease: A systematic review. Bioengineering (Basel) 2022; 9(1): 27. http://dx.doi.org/10.3390/bioengineering9010027 PMID: 35049736

[10] Hoffmann I, Nemeth D, Dye CD, Pákáski M, Irinyi T, Kálmán J. Temporal parameters of spontaneous speech in Alzheimer's disease. Int J Speech Lang Pathol 2010; 12(1): 29-34. http://dx.doi.org/10.3109/17549500903137256 PMID: 20380247

[11] Boersma P. Praat, a system for doing phonetics by computer. Glot Int 2002; 5: 341-5.

[12] Szatloczki G, Hoffmann I, Vincze V, Kalman J, Pakaski M. Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. Front Aging Neurosci 2015; 7: 195-202. http://dx.doi.org/10.3389/fnagi.2015.00195 PMID: 26539107

[13] Tóth L, Gosztolya G, Vincze V, Hoffmann I, Szatlóczki G, Biró E, *et al.* Automatic detection of mild cognitive impairment from spontaneous speech using ASR. Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech). 2694-8. http://dx.doi.org/10.21437/Interspeech.2015-568

[14] Gosztolya G, Tóth L, Grósz T, *et al.* Detecting mild cognitive impairment from spontaneous speech by correlation-based phonetic feature selection. Proceedings of Interspeech 2016; 107-111: 107-11. http://dx.doi.org/10.21437/Interspeech.2016-384

[15] Vincze V, Gosztolya G, Tóth L, Hoffmann I, Szatlóczki G, Bánréti Z, *et al.* Detecting mild cognitive impairment by exploiting linguistic information from transcripts. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL). 181-7. http://dx.doi.org/10.18653/v1/P16-2030

[16] Tóth L, Hoffmann I, Gosztolya G, *et al.* A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. Curr Alzheimer Res 2018; 15(2): 130-8. http://dx.doi.org/10.2174/1567205014666171121114930 PMID: 29165085

[17] Gosztolya G, Vincze V, Tóth L, Pákáski M, Kálmán J, Hoffmann I. Identifying mild cognitive impairment and mild Alzheimer's

disease based on spontaneous speech using ASR and linguistic features. Comput Speech Lang 2019; 53: 181-97.
http://dx.doi.org/10.1016/j.csl.2018.07.007

[18] Solomon A, Mangialasche F, Richard E, *et al.* Advances in the prevention of Alzheimer's disease and dementia. J Intern Med 2014; 275(3): 229-50.
http://dx.doi.org/10.1111/joim.12178 PMID: 24605807

[19] de la Fuente Garcia S, Ritchie CW, Luz S. Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: A systematic review. J Alzheimers Dis 2020; 78(4): 1547-74.
http://dx.doi.org/10.3233/JAD-200888 PMID: 33185605

[20] Chien Y, Hong S, Cheah W, Fu L, Chang Y. An assessment system for Alzheimer's disease based on speech using a novel feature sequence design and recurrent neural network. IEEE International Conference on Systems, Man, and Cybernetics (SMC). 3289-94.
http://dx.doi.org/10.1109/SMC.2018.00557

[21] Roark B, Mitchell M, Hosom JP, Hollingshead K, Kaye J. Spoken language derived measures for detecting mild cognitive impairment. IEEE Trans Audio Speech Lang Process 2011; 19(7): 2081-90.
http://dx.doi.org/10.1109/TASL.2011.2112351 PMID: 22199464

[22] Sajjadi SA, Patterson K, Tomek M, Nestor PJ. Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. Aphasiology 2012; 26(6): 847-66.
http://dx.doi.org/10.1080/02687038.2012.654933

[23] Jarrold W, Peintner B, Wilkins D, Vergryi D, Richey C, Gorno-Tempini ML, *et al.* Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. Proceedings of the Workshop on Computational Linguistics and Clinical Psychology (CLPsych). 27-37.
http://dx.doi.org/10.3115/v1/W14-3204

[24] De Looze C, Kelly F, Crosby L, *et al.* Changes in speech chunking in reading aloud is a marker of mild cognitive impairment and mild-to-moderate Alzheimer's disease. Curr Alzheimer Res 2018; 15(9): 828-47.
http://dx.doi.org/10.2174/1567205015666180404165017 PMID: 29623841

[25] Luz S, de la Fuente S, Albert P. A method for analysis of patient speech in dialogue for dementia detection. arXiv 2018; 1811.09919.

[26] Guo Z, Ling Z, Li Y. Detecting Alzheimer's disease from continuous speech using language models. J Alzheimers Dis 2019; 70(4): 1163-74.
http://dx.doi.org/10.3233/JAD-190452 PMID: 31322577

[27] Sluis RA, Angus D, Wiles J, *et al.* An automated approach to examining pausing in the speech of people with dementia. Am J Alzheimers Dis Other Demen 2020; 35: 1533317520939773.
http://dx.doi.org/10.1177/1533317520939773 PMID: 32648470

[28] König A, Satt A, Sorin A, *et al.* Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement (Amst) 2015; 1(1): 112-24.
http://dx.doi.org/10.1016/j.dadm.2014.11.012 PMID: 27239498

[29] König A, Satt A, Sorin A, *et al.* Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. Curr Alzheimer Res 2018; 15(2): 120-9.
http://dx.doi.org/10.2174/1567205014666170829111942 PMID: 28847279

[30] Mirzaei S, El Yacoubi M, Garcia-Salicetti S, *et al.* Two-stage feature selection of voice parameters for early Alzheimer's disease prediction. IRBM 2018; 39(6): 430-5.
http://dx.doi.org/10.1016/j.irbm.2018.10.016

[31] Tröger J, Linz N, König A, *et al.* Exploitation *vs.* exploration-computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. Neuropsychologia 2019; 131: 53-61.
http://dx.doi.org/10.1016/j.neuropsychologia.2019.05.007 PMID: 31121184

[32] Satt A, Sorin A, Toledo-Ronen O, Barkan O, Kompatsiaris I, Kokonozi A, *et al.* Evaluation of speech-based protocol for detection of early-stage dementia. Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech). 1692-6.
http://dx.doi.org/10.21437/Interspeech.2013-32

[33] Gosztolya G, Balogh R, Imre N, *et al.* Cross-lingual detection of mild cognitive impairment based on temporal parameters of spontaneous speech. Comput Speech Lang 2021; 69: 101215.
http://dx.doi.org/10.1016/j.csl.2021.101215

[34] Balogh R, Imre N, Gosztolya G, Hoffmann L, Pákáski M, Kálmán J. The role of silence in verbal fluency tasks – a new approach for the detection of mild cognitive impairment. J Int Neuropsychol Soc 2022; 1-13.
http://dx.doi.org/10.1017/S1355617721001454 PMID: 35067261

[35] Beltrami D, Gagliardi G, Rossini Favretti R, Ghidoni E, Tamburini F, Calzà L. Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? Front Aging Neurosci 2018; 10: 369.
http://dx.doi.org/10.3389/fnagi.2018.00369

[36] Kato S, Endo H, Homma A, Sakuma T, Watanabe K. Early detection of cognitive impairment in the elderly based on Bayesian mining using speech prosody and cerebral blood flow activation. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 5813-6.
http://dx.doi.org/10.1109/EMBC.2013.6610873

[37] Tanaka H, Adachi H, Ukita N, *et al.* Detecting dementia through interactive computer avatars. IEEE J Transl Eng Health Med 2017; 5: 2200111.
http://dx.doi.org/10.1109/JTEHM.2017.2752152 PMID: 29018636

[38] Kobayashi M, Kosugi A, Takagi H, Nemoto M, Nemoto K, Arai T, *et al.* Effects of age-related cognitive decline on elderly user interactions with voice-based dialogue systems. IFIP Conference on Human-Computer Interaction. 53-74.
http://dx.doi.org/10.1007/978-3-030-29390-1_4

[39] Yamada Y, Shinkawa K, Kobayashi M, *et al.* Tablet-based automatic assessment for early detection of Alzheimer's disease using speech responses to daily life questions. Front Digit Health 2021; 3: 653904.
http://dx.doi.org/10.3389/fdgth.2021.653904 PMID: 34713127

[40] Nasrolahzadeh M, Mohammadpoory Z, Haddadnia J. Higher-order spectral analysis of spontaneous speech signals in Alzheimer's disease. Cogn Neurodynamics 2018; 12(6): 583-96.
http://dx.doi.org/10.1007/s11571-018-9499-8 PMID: 30483366

[41] Meilán JJG, Martínez-Sánchez F, Carro J, López DE, Millian-Morell L, Arana JM. Speech in Alzheimer's disease: Can temporal and acoustic parameters discriminate dementia? Dement Geriatr Cogn Disord 2014; 37(5-6): 327-34.
http://dx.doi.org/10.1159/000356726 PMID: 24481220

[42] López-de-Ipiña K, Alonso JB, Travieso CM, *et al.* On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis. Sensors (Basel) 2013; 13(5): 6730-45.
http://dx.doi.org/10.3390/s130506730 PMID: 23698268

[43] Espinoza-Cuadros F, Garcia-Zamora MA, Torres-Boza D, Ferrer-Riesgo CA, Montero-Benavides A, Gonzales-Moreira E, *et al.* A spoken language database for research on moderate cognitive impairment: Design and preliminary analysis. Lect Notes Comput Sci 2014; 8854: 219-28. [IberSpeech].
http://dx.doi.org/10.1007/978-3-319-13623-3_23

[44] Gonzalez-Moreira E, Torres-Boza D, Kairuz HA, *et al.* Automatic prosodic analysis to identify mild dementia. BioMed Res Int 2015; 2015: 916356.
http://dx.doi.org/10.1155/2015/916356 PMID: 26558287

[45] Fraser KC, Lundholm Fors K, Eckerström M, Öhman F, Kokkinakis D. Predicting MCI status from multimodal language data using cascaded classifiers. Front Aging Neurosci 2019; 11: 205.
http://dx.doi.org/10.3389/fnagi.2019.00205 PMID: 31427959

[46] Themistocleous C, Eckerström M, Kokkinakis D. Voice quality and speech fluency distinguish individuals with Mild Cognitive Impairment from Healthy Controls. PLoS One 2020; 15(7): e0236009.
http://dx.doi.org/10.1371/journal.pone.0236009 PMID: 32658934

[47] Khodabakhsh A, Demiroglu C. Analysis of speech-based measures for detecting and monitoring Alzheimer's disease. Methods Mol Biol 2015; 1246: 159-73.
http://dx.doi.org/10.1007/978-1-4939-1985-7_11 PMID: 25417086

[48] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. Arch Neurol 1999; 56(3): 303-8. http://dx.doi.org/10.1001/archneur.56.3.303 PMID: 10190820
[49] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12(3): 189-98. http://dx.doi.org/10.1016/0022-3956(75)90026-6 PMID: 1202204
[50] Manos PJ, Wu R. The ten point clock test: A quick screen and grading method for cognitive impairment in medical and surgical patients. Int J Psychiatry Med 1994; 24(3): 229-44. http://dx.doi.org/10.2190/5A0F-936P-VG8N-0F5R PMID: 7890481
[51] Yesavage JA, Brink TL, Rose TL, *et al.* Development and validation of a geriatric depression screening scale: A preliminary report. J Psychiatr Res 1982-1983; 17(1): 37-49. http://dx.doi.org/10.1016/0022-3956(82)90033-4 PMID: 7183759
[52] Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS) – Recent evidence and development of a shorter version. Clin Gerontol 1986; 5: 165-73. http://dx.doi.org/10.1300/J018v05n01_09
[53] Young S, Evermann G, Gales MJF, Hain T, Kershaw D, Liu X, *et al.* The HTK Book. Cambridge, UK: Cambridge University 2006.
[54] Hinton G, Deng L, Yu D, *et al.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Process Mag 2012; 29(6): 82-97. http://dx.doi.org/10.1109/MSP.2012.2205597
[55] Tóth L. Phone recognition with hierarchical convolutional deep maxout networks. EURASIP J Audio Speech Music Process 2015; 25(1): 1-3. http://dx.doi.org/10.1186/s13636-015-0068-3
[56] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS). 315-23.
[57] Rousseau A, Delglise P, Estve Y. TED-LIUM: An automatic speech recognition dedicated corpus. Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12). 125-9.
[58] Neuberger T, Gyarmathy D, Gráczi TE, Horváth V, Gósy M, Beke A. Development of a large spontaneous speech database of agglutinative Hungarian language. Text Speech and Dialogue 2014; pp. 424-31. http://dx.doi.org/10.1007/978-3-319-10816-2_51
[59] Eklund R. Prolongations: A dark horse in the disfluency stable. Proceedings of Disfluency in Spontaneous Speech (DiSS) 2001; 5-8.
[60] Deme A, Markó A. Lengthenings and filled pauses in Hungarian adults' and children's speech. Proceedings of Disfluency in Spontaneous Speech (DiSS) 2013; 21-4.
[61] Gayraud F, Lee HR, Barkat-Defradas M. Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. Clin Linguist Phon 2011; 25(3): 198-209. http://dx.doi.org/10.3109/02699206.2010.521612 PMID: 21080826
[62] Lee H, Gayraud F, Hirsh F, Barkat-Defradas M. Speech dysfluencies in normal and pathological aging: A comparison between Alzheimer patients and healthy elderly subjects. Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS). 1174-7.
[63] Singh S, Bucks RS, Cuerden JM. Evaluation of an objective technique for analysing temporal variables in DAT spontaneous speech. Aphasiology 2001; 15(6): 571-83. http://dx.doi.org/10.1080/02687040143000041
[64] Sarazin M, Chauviré V, Gerardin E, *et al.* The amnestic syndrome of hippocampal type in Alzheimer's disease: An MRI study. J Alzheimers Dis 2010; 22(1): 285-94. http://dx.doi.org/10.3233/JAD-2010-091150 PMID: 20847406
[65] Pistono A, Jucla M, Barbeau EJ, *et al.* Pauses during autobiographical discourse reflect episodic memory processes in early Alzheimer's disease. J Alzheimers Dis 2016; 50(3): 687-98. http://dx.doi.org/10.3233/JAD-150408 PMID: 26757034
[66] Simons JS, Gilbert SJ, Owen AM, Fletcher PC, Burgess PW. Distinct roles for lateral and medial anterior prefrontal cortex in contextual recollection. J Neurophysiol 2005; 94(1): 813-20. http://dx.doi.org/10.1152/jn.01200.2004 PMID: 15728761
[67] Roca M, Torralva T, Gleichgerrcht E, *et al.* The role of Area 10 (BA10) in human multitasking and in social cognition: A lesion study. Neuropsychologia 2011; 49(13): 3525-31. http://dx.doi.org/10.1016/j.neuropsychologia.2011.09.003 PMID: 21930138
[68] Arlt S, Buchert R, Spies L, Eichenlaub M, Lehmbeck JT, Jahn H. Association between fully automated MRI-based volumetry of different brain regions and neuropsychological test performance in patients with amnestic mild cognitive impairment and Alzheimer's disease. Eur Arch Psychiatry Clin Neurosci 2013; 263(4): 335-44. http://dx.doi.org/10.1007/s00406-012-0350-7 PMID: 22940716
[69] Wilson SM, Henry ML, Besbris M, *et al.* Connected speech production in three variants of primary progressive aphasia. Brain 2010; 133(Pt 7): 2069-88. http://dx.doi.org/10.1093/brain/awq129 PMID: 20542982
[70] Hernández-Domínguez L, Ratté S, Sierra-Martínez G, Roche-Bergua A. Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. Alzheimers Dement (Amst) 2018; 10(1): 260-8. http://dx.doi.org/10.1016/j.dadm.2018.02.004 PMID: 29780871
[71] Gauthier S, Reisberg B, Zaudig M, *et al.* Mild cognitive impairment. Lancet 2006; 367(9518): 1262-70. http://dx.doi.org/10.1016/S0140-6736(06)68542-5 PMID: 16631882