Detecting Mild Cognitive Impairment
by Exploiting Linguistic Information from Transcripts

Veronika Vincze[1,2], **Gábor Gosztolya**[1,2], **László Tóth**[1], **Ildikó Hoffmann**[3,4],

Gréta Szatlóczki[5], **Zoltán Bánréti**[4], **Magdolna Pákáski**[5] and **János Kálmán**[5]

[1] MTA-SZTE Research Group for Artificial Intelligence
[2] University of Szeged, Institute of Informatics
[3] University of Szeged, Department of Linguistics
[4] Research Institute of Linguistics, Hungarian Academy of Sciences
[5] University of Szeged, Department of Psychiatry
`vinczev@inf.u-szeged.hu`

Abstract

Here we seek to automatically identify Hungarian patients suffering from mild cognitive impairment (MCI) based on linguistic features collected from their speech transcripts. Our system uses machine learning techniques and is based on several linguistic features like characteristics of spontaneous speech as well as features exploiting morphological and syntactic parsing. Our results suggest that it is primarily morphological and speech-based features that help distinguish MCI patients from healthy controls.

1 Background

Mild cognitive impairment (MCI) is a heterogeneous set of symptoms that are essential in the early detection of Alzheimer's Disease (AD) (Negash et al., 2007). Symptoms such as language dysfunctions may occur even nine years before the actual diagnosis (APA, 2000). Thus, the language use of the patient may often indicate MCI well before the clinical diagnosis of dementia.

MCI is known to influence the (spontaneous) speech of the patient via three main aspects. First, verbal fluency declines, which results in longer hesitations and a lower speech rate (Roark et al., 2011). Second, the lexical frequency of words and part-of-speech tags may also change significantly as the patient has problems with finding words (Croot et al., 2000). Third, the emotional responsiveness of the patient was also observed to change in many cases (Lopez-de Ipiña et al., 2015).

For many patients, MCI is never recognized as in the early stage of the disease it is not trivial even for experts to detect cognitive impairment: according to Boise et al. (2004), up to 50% of MCI patients are never diagnosed with MCI. Although there are well known tests such as the Mini Mental Test, they are usually not sensitive enough to reliably filter out MCI in its early stage. Tests on linguistic memory prove more efficient in detecting MCI, but they tend to yield a relatively high number of false positive diagnoses (Roark et al., 2011).

Although language abilities are impaired from an early stage of the disease, evaluating the language capacities of the patients has only received marginal attention when diagnosing AD (Bayles, 1982). However, if diagnosed early, a proper medical treatment may delay the occurrence of other (more severe) symptoms of dementia to the latest extent possible (Kálmán et al., 2013).

Here we seek to automatically identify Hungarian patients suffering from mild cognitive impairment based on their speech transcripts. Our system uses machine learning techniques and is based on several features like linguistic characteristics of spontaneous speech as well as features exploiting morphological and syntactic parsing.

Recently, several studies have reported results on identifying different types of dementia with NLP and speech recognition techniques. For instance, automatic speech recognition tools were employed in detecting aphasia (Fraser et al., 2013b; Fraser et al., 2014; Fraser et al., 2013a) and mild cognitive impairment (Lehr et al., 2012), and Alzheimer's Disease (Baldas et al., 2010; Satt et al., 2014). Jarrold et al. (2014) distinguished four types of dementia on the basis of spontaneous speech samples. Lexical analysis of spontaneous speech may also indicate different types of dementia (Bucks et al., 2000; Holmes and Singh, 1996) and may be exploited in the automatic detection of patients suffering from dementia (Thomas et al.,

2005). As for analyzing written language, changes in the writing style of people may also refer to dementia (Garrard et al., 2005; Hirst and Wei Feng, 2012; Le et al., 2011).

Concerning the automatic detection of MCI in Hungarian subjects, Tóth et al. (2015) experimented with speech recognition techniques. However, to the best of our knowledge, this is the first attempt to identify MCI on the basis of written texts, i.e. speech transcripts for Hungarian.

In the long run, we would like to develop a system that can automatically detect linguistic symptoms of MCI in its early stage, so that the person can get medical treatment as early as possible. It should be noted, however, that our goal cannot be an official diagnosis as diagnosing patients requires medical experience. All we can do is implement a test supported by methods used in artificial intelligence, which indicates whether the patient is at risk and if so, s/he can turn to medical experts who will provide the clinical diagnosis.

2 Data

In our experiments[1], two short animated films were presented to the patients at the memory ambulance of the University of Szeged. Patients were asked to talk about the first film then about their previous day, and lastly, about the second film. Their speech productions were recorded and transcribed by linguists, who explicitly marked speech phenomena like hesitations and pauses in the transcripts. These transcripts formed the basis of our experiments, i.e. we exploited only written information.

All of our 84 subjects were native speakers of Hungarian, a morphologically rich language. For each person, a clinical diagnosis was at our disposal, i.e. it was clinically proved whether the patient suffers from MCI or not. On the basis of these data, subjects were classified as either MCI patient or healthy control at the university memorial. Table 1 shows data on the subjects' gender and diagnosis while Table 2 shows the mean values for age and education (in terms of years attended at school).

Speech transcripts reflect several characteristics of spontaneous speech. On the one hand, they contain several forms of hesitations and silent pauses,

	MCI	Control	Total
Male	16	13	29
Female	32	23	55
Total	48	36	84

Table 1: Subjects' gender and diagnosis.

	MCI	Control	Significance
age	73.08	69.28	0.0124
education	11.42	12.47	

Table 2: Mean values of demographic features and level of statistical significance in terms of p-value.

which are also marked in the transcripts, on the other hand, they abound in phenomena typical of spontaneous Hungarian speech such as phonological deletion (*mer* instead of the standard form *mert* "because" or *ement* instead of the standard form *elment* "(he) left") and lengthening (*utánna* instead of the standard form *utána* "then"). There are duplications (*ez ezt* "this this-ACC") and neologisms created by the speaker (*feltkáva*, which probably means *főtt kávé* "boiled coffee").

Fillers also deserve special attention when studying transcripts. Besides hesitations, we treated words and phrases referring to some kind of uncertainty together with indefinite pronouns as fillers such as *ilyen* "such", *olyan* "such", *izé* "thing, gadget", *és aztán* "and then", *valamilyen* "some kind of", *valahogy* "somehow", *valamerre* "somewhere"[2]. Thus, MCI patients often seem to substitute content words with fillers or indefinite pronouns, moreover, they also appear to use lots of paraphrases, which also indicate uncertainty just like *egy ilyen bagolyszerűség* a such owl-likeness "something similar to an owl" or *az olyan délelőtt volt* that such morning was "that happened some time in the morning".

3 Experiments

In order to determine the status of the subjects, we experimented with machine learning tools. The task was regarded as binary classification, i.e. subjects were classified as either an MCI patient or a healthy control, on the basis of a feature set derived from their transcripts.

At first, transcripts were morphologically and syntactically analysed with magyarlanc, a linguistic preprocessing toolkit developed for Hungarian

---

[1]Our experiments conform to all operative rules and restrictions on data collection, anonymization and publication according to the requirements in the European Union.

[2]This words seem to have a lot in common with weasel and hedge words, which refer to uncertainty (Vincze, 2013).

(Zsibrita et al., 2013). For classification, we exploited morphological, syntactic and semantic features extracted from the output of magyarlanc.

Each person was asked to recall three different stories. As MCI is strongly related to memory deficit, we believe that the order of the tasks might also influence performance, hence we opted for processing each transcript separately. Thus, for each person, features to be discussed below were calculated separately for the three transcripts and all of them were exploited in the system.

3.1 Feature set

In our experiments, we employed features of spontaneous speech and morphological and semantic features derived from the transcripts and their automatic linguistic analyses. When defining our features, we took into account the fact that the speech of MCI patients may contain more pauses and hesitations than that of healthy controls (Tóth et al., 2015) and they are also supposed to have a restricted vocabulary due to cognitive deficit, which may affect the choice of words and the frequency of parts of speech (Croot et al., 2000) and might even yield neologisms. We also made use of demographic features that were at our disposal.

Our feature set contained the following features:

Spontaneous speech based features:

number of filled and silent pauses; number and rate of hesitations compared to the number of tokens; number of pauses that follow an article and precede content words as this might reflect that MCI patients may have difficulties with finding the appropriate content words; number of lengthened sounds (which we considered as a special form of hesitation).

Morphological features:

number of tokens and words; number and rate of distinct lemmas; number of punctuation marks; number and rate of nouns, verbs, adjectives, pronouns and conjunctions; number of first person singular verbs as it might also be indicative how often the patient reflects to him/herself; number and rate of unanalyzed words, i.e. those with an "unknown" POS tag, which might indicate neologisms created by the speaker on the spot.

Semantic features:

number and rate of fillers and uncertain words compared to the number of all tokens; number and rate of words/phrases related to memory activity (e.g. *nem emlékszem* not remember-1SG "I can't remember") as they directly signal problems with memory and recall; number of negation words; number and rate of content words and function words; number of thematic words related to the content of the films, based on manually constructed lists.

Demographic features:

gender; age; education.

The mean values for each feature are reported in Table 3.

3.2 Statistical analysis of features

In order to reveal which features can most effectively distinguish healthy controls from MCI patients, we carried out a statistical analysis of the data (t-tests for each feature and transcript). For most of the features, significant differences were found between the two groups – p-values are listed in Table 3. The age of the patients also indicates significant differences: people who were at least 71 years old were more probable to suffer from MCI than those who were younger at the time of the experiment ($p = 0.0124$).

According to the data, each group of features has a significant effect in distinguishing controls and MCI patients. It is shown that it is mostly the second transcript (the one including the narratives about the subjects' previous days) where significant differences may be found among MCI patients and the control group. However, significant differences exist for the other two types of texts as well.

3.3 Machine learning experiments

To automatically identify MCI patients, we exploited machine learning techniques, i.e. support vector machines (SVM) (Cortes and Vapnik, 1995) with the default settings of Weka (Hall et al., 2009) and due to the small size of the dataset, we applied leave-one-out cross validation. As a baseline, majority labeling was used. For the evaluation, the accuracy, precision, recall and F-measure metrics were utilized.

In order to examine the effect of certain groups of features, we carried out an ablation study, i.e. we retrained the system without making use of one specific group of features. The results and differences are shown in Table 4.

Feature	T1		T2		T3		Significance		
	MCI	control	MCI	control	MCI	control	T1	T2	T3
token #	141.46	126.31	209.40	129.72	124.21	121.33		0.0017	
sentence #	8.38	8.50	13.42	10.22	8.08	8.11			
token %	22.40	18.42	20.01	16.04	20.19	19.25		0.0015	
word #	115.54	101.53	168.65	101.39	100.27	99.36			
lemma #	68.40	64.31	101.19	70.86	61.85	61.50		0.0017	
lemma %	0.51	0.54	0.54	0.59	0.53	0.53		0.0214	
verb #	21.71	21.31	36.63	24.11	19.56	19.97		0.0033	
verb %	0.19	0.21	0.23	0.25	0.20	0.21	0.0009		
noun #	23.98	25.42	33.23	23.00	21.69	21.97		0.0149	
noun %	0.21	0.25	0.19	0.24	0.22	0.22	0.0004	0.0001	
adjective #	6.13	3.75	9.50	5.47	4.77	5.50	0.0068	0.0051	
adjective %	0.05	0.04	0.06	0.05	0.04	0.05	0.0067	0.0259	
pronoun #	14.29	10.11	15.85	7.67	14.21	13.25	0.0082	0.0001	
pronoun %	0.12	0.10	0.09	0.08	0.14	0.13	0.0053	0.0227	
conjunction #	12.69	9.53	18.19	8.72	10.81	10.14	0.0345	0.0009	
conjunction %	0.10	0.09	0.10	0.08	0.10	0.10		0.0417	
Sg1 verb #	3.42	2.25	18.94	13.36	2.63	2.64	0.0341	0.0224	
punctuation #	25.92	24.78	40.75	28.33	23.94	21.97		0.0062	
unknown word #	0.31	0.19	0.31	0.11	0.08	0.08			
unknown word %	0.21	0.25	0.12	0.10	0.07	0.07			
filled pause #	3.65	2.44	3.92	1.56	2.35	1.44		0.0319	
pause #	12.63	9.11	19.77	9.89	11.15	7.28		0.0008	0.0191
pause after article #	1.40	1.08	1.23	0.72	1.29	0.81			0.0449
lengthened sound #	24.35	20.39	35.44	19.89	19.94	18.22		0.0008	
hesitation #	17.40	12.39	25.92	12.25	14.71	9.25	0.0362	0.0007	0.0047
hesitation %	12.93	9.32	12.67	10.19	12.06	7.55	0.0216		0.0010
uncertain word #	6.44	4.83	7.48	2.81	6.23	5.89		0.0003	
uncertain word %	4.36	3.69	3.15	2.07	4.89	4.89		0.0087	
memory word #	1.23	0.69	0.54	0.17	0.96	1.14	0.0211	0.0166	
memory word %	0.93	0.56	0.28	0.12	0.72	0.83			
film word 1	10.56	12.92	0.21	0.14	4.02	4.75	0.0325		
film word 2	5.75	5.69	0.33	0.28	9.10	11.06			0.0291
content word %	0.60	0.63	0.69	0.72	0.60	0.61	0.0441	0.0042	
function word %	0.39	0.37	0.31	0.28	0.40	0.39	0.0342	0.0041	
negation #	2.42	1.39	3.50	1.47	2.13	2.17	0.0305	0.0034	

Table 3: Mean values of features and level of statistical significance in terms of p-value. #: number, %:ratio, T: transcript.

Features	MCI			Control			Total			%
	P	R	F	P	R	F	P	R	F	
all included	72.0	75.0	73.5	64.7	61.1	62.9	68.9	69.0	68.9	69.1
w/o semantic	75.0	81.3	78.0	71.9	63.9	67.6	73.7	73.8	73.6	73.8
	+3.0	+6.3	+4.5	+7.2	+2.8	+4.7	+4.8	+4.8	+4.7	+4.7
w/o demographic	70.0	72.9	71.4	61.8	58.3	60.0	66.5	66.7	66.5	66.7
	-2.0	-2.1	-2.1	-2.9	-2.8	-2.9	-2.4	-2.3	-2.4	-2.4
w/o speech-based	70.8	70.8	70.8	61.1	61.1	61.1	66.7	66.7	66.7	66.7
	-1.2	-4.2	-2.7	-3.6	0.0	-1.8	-2.2	-2.3	-2.2	-2.4
w/o morphological	72.3	70.8	71.6	62.2	63.9	63.0	68.0	67.9	67.9	67.9
	+0.3	-4.2	-1.9	-2.5	+2.8	+0.1	-0.9	-1.1	-1.0	-1.2
only significant	81.4	72.9	76.9	68.3	77.8	72.7	75.8	75.0	75.1	75.0
	+9.4	-2.1	+3.4	+3.6	+16.7	+9.8	+6.9	+6.0	+6.2	+5.9

Table 4: Results and differences. MCI: mild cognitive impairment, P: precision, R: recall, F: F-measure, %: accuracy.

4 Results and Discussion

Using all the features, our system managed to achieve an accuracy score of 69.1%, that is, 58 out of the 84 patients were correctly diagnosed. 12 patients were falsely diagnosed as healthy and 14 controls were falsely labeled as MCI patients. Our results outperformed the baseline (57.14% in terms of accuracy). The system got a high recall value for MCI patients (75.0) but a lower one for controls (61.1), which is encouraging in the light of the fact that our main goal is to identify the widest possible range of potential MCI patients, who can turn to clinical experts to find out what their clinical diagnosis is.

We also experimented with using only features that displayed statistically significant differences among controls and MCI patients (see Table 3). Somewhat surprisingly, an accuracy of 75% could be achieved in this way, which indicates that some of our original features are superfluous and just confused the system, and this result needs further investigation.

An ablation study was also carried out to analyze the added value of each feature group. Speech-based, demographic and morphological features unequivocally contributed to performance. However, the effect of semantic features seems less obvious as they harm performance taken as a whole but some individual semantic features are useful for the system, as shown by the results achieved with just using significant features.

When investigating the errors made by our system, we found that MCI patients that spoke only a few short sentences were often classified as healthy controls. They had a lower number and rate of hesitations and pauses, moreover, their vocabulary contained fewer fillers and uncertain words, and these features resemble those typical of healthy controls. What is more, healthy subjects who talked more also hesitated more, which might be indicative of MCI. Furthermore, their use of pronouns and conjunctions was also more similar to those of MCI patients, hence the system falsely predicted a positive diagnosis for them.

Due to the specific characteristics of the data and the complexity of data collection – which requires clinical experiments – our dataset can be expanded only step by step. However, we found statistically significant differences among MCI patients and healthy controls concerning several linguistic and speech-based features even in our small dataset, which may be beneficial for our future experiments and might be also exploited by those who study spontaneous speech.

5 Conclusions

In this study, we introduced our system that automatically detects Hungarian patients suffering from mild cognitive impairment on the basis of their speech transcripts. The system is based on features derived from morphological and syntactic analysis as well as characteristics of spontaneous speech. Both statistical and machine learning results revealed that morphological and spontaneous speech-based features have an essential role in distinguishing MCI patients from healthy controls.

In the future, we would like to extend our dataset with new transcripts. Also, we intend to improve our machine learning system and investigate the role of semantic features. Lastly, we would like to integrate features from automatic speech recognition into our system so that tools from both speech technology and natural language processing can contribute to the automatic detection of mild cognitive impairment.

References

APA. 2000. *DSM-IV-TR*. American Psychiatric Association.

Vassilis Baldas, Charalampos Lampiris, Christos N. Capsalis, and Dimitrios Koutsouris. 2010. Early Diagnosis of Alzheimer's Type Dementia Using Continuous Speech Recognition. In James C. Lin and Konstantina S. Nikita, editors, *MobiHealth*, volume 55 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 105–110. Springer.

Kathryn A Bayles. 1982. Language function in senile dementia. *Brain and Language*, 16(2):265–280.

Linda Boise, Margaret B Neal, and Jeffrey Kaye. 2004. Dementia assessment in primary care: Results from a study in three managed care systems. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(6):M621–M626.

R.S. Bucks, S. Singh, J.M. Cuerden, and G.K. Wilcock. 2000. Analysis of spontaneous, conversational speech in dementia of alzheimer type: evaluation of an objective technique for analysing lexical performance. *Aphasiology*, 14(1):71–91.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20(3):273–297.

Karen Croot, John R. Hodges, John Xuereb, and Karalyn Patterson. 2000. Phonological and articulatory impairment in alzheimer's disease: A case series. *Brain and Language*, 75(2):277 – 309.

Kathleen Fraser, Frank Rudzicz, Naida Graham, and Elizabeth Rochon. 2013a. Automatic speech recognition in the diagnosis of primary progressive aphasia. *Proceedings of the Fourth Workshop on Speech and Language Processing for Assistive Technologies*, pages 47–54.

Kathleen C. Fraser, Frank Rudzicz, and Elizabeth Rochon. 2013b. Using text and acoustic features to diagnose progressive aphasia and its subtypes. In Frdric Bimbot, Christophe Cerisara, Ccile Fougeron, Guillaume Gravier, Lori Lamel, Franois Pellegrino, and Pascal Perrier, editors, *INTERSPEECH*, pages 2177–2181. ISCA.

Kathleen C Fraser, Jed A Meltzer, Naida L Graham, Carol Leonard, Graeme Hirst, Sandra E Black, and Elizabeth Rochon. 2014. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex*, 55:43–60.

Peter Garrard, Lisa M Maloney, John R Hodges, and Karalyn Patterson. 2005. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, 128(2):250–260.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18.

Graeme Hirst and Vanessa Wei Feng. 2012. Changes in style in authors with Alzheimer's disease. *English Studies*, 93(3):357–370.

David I. Holmes and Sameer Singh. 1996. A stylometric analysis of conversational speech of aphasic patients. *Literary and Linguistic Computing*, 11:133–140.

William Jarrold, Bart Peintner, David Wilkins, Dimitra Vergryi, Colleen Richey, Maria L. Gorno-Tempini, and Jennifer Ogar. 2014. Aided diagnosis of dementia type through computer-based analysis of spontaneous speech. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 27–37, Baltimore, Maryland, USA, June. Association for Computational Linguistics.

János Kálmán, Magdolna Pákáski, Ildikó Hoffmann, Gergely Drótos, Gyöngyi Darvas, Krisztina Boda, Tamás Bencsik, Aliz Gyimesi, Zsófia Gulyás, Magdolna Bálint, Gréta Szatlóczki, and Edina Papp. 2013. Early mental test – developing a screening test for mild cognitive impairment. *Ideggyógyászati szemle*, 66(1-2):43–52.

Xuan Le, Ian Lancashire, Graeme Hirst, and Regina Jokel. 2011. Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. *Literary and Linguistic Computing*, 26(4):435–461.

Maider Lehr, Emily Tucker Prud'hommeaux, Izhak Shafran, and Brian Roark. 2012. Fully automated neuropsychological assessment for detecting mild cognitive impairment. In *INTERSPEECH*, pages 1039–1042. ISCA.

Karmele Lopez-de Ipiña, Jesús B Alonso, Jordi Solé-Casals, Nora Barroso, Patricia Henriquez, Marcos Faundez-Zanuy, Carlos M Travieso, Miriam Ecay-Torres, Pablo Martinez-Lage, and Harkaitz Eguiraun. 2015. On automatic diagnosis of alzheimers disease based on spontaneous speech analysis and emotional temperature. *Cognitive Computation*, 7(1):44–55.

Selam Negash, Lindsay E Petersen, Yonas E Geda, David S Knopman, Bradley F Boeve, Glenn E Smith, Robert J Ivnik, Darlene V Howard, James H Howard Jr, and Ronald C Petersen. 2007. Effects of ApoE genotype and Mild Cognitive Impairment on implicit learning. *Neurobiology of Aging*, 28(6):885–893.

Brian Roark, Margaret Mitchell, John-Paul Hosom, Kristy Hollingshead, and Jeffrey Kaye. 2011. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7):2081–2090.

Aharon Satt, Ron Hoory, Alexandra König, Pauline
Aalten, and Philippe H. Robert. 2014. Speech-
based automatic and robust detection of very early
dementia. In *15th Annual Conference of the Inter-
national Speech Communication Association*, pages
2538–2542.

Calvin Thomas, Vlado Kešelj, Nick Cercone, Kenneth
Rockwood, and Elissa Asp. 2005. Automatic de-
tection and rating of dementia of Alzheimer type
through lexical analysis of spontaneous speech. In
*Mechatronics and Automation, 2005 IEEE Inter-
national Conference*, volume 3, pages 1569–1574.
IEEE.

László Tóth, Gábor Gosztolya, Veronika Vincze, Ildikó
Hoffmann, Gréta Szatlóczki, Edit Biró, Fruzsina
Zsura, Magdolna Pákáski, and János Kálmán. 2015.
Automatic detection of mild cognitive impairment
from spontaneous speech using ASR. In *16th An-
nual Conference of the International Speech Com-
munication Association*, pages 2694–2698.

Veronika Vincze. 2013. Weasels, Hedges and Pea-
cocks: Discourse-level Uncertainty in Wikipedia
Articles. In *Proceedings of the Sixth International
Joint Conference on Natural Language Processing*,
pages 383–391, Nagoya, Japan, October. Asian Fed-
eration of Natural Language Processing.

János Zsibrita, Veronika Vincze, and Richárd Farkas.
2013. magyarlanc: A toolkit for morphological and
dependency parsing of Hungarian. In *Proceedings
of RANLP*, pages 763–771.