

## RESEARCH ARTICLE

# Use of Speech Analyses within a Mobile Application for the Assessment of Cognitive Impairment in Elderly People

A. König<sup>1,3,\*</sup>, A. Satt<sup>2</sup>, A. Sorin<sup>2</sup>, R. Hoory<sup>2</sup>, A. Derreumaux<sup>1</sup>, R. David<sup>1,3,4</sup> and P.H. Robert<sup>1,3,4</sup>

<sup>1</sup>EA CoBTek, University of Nice Sophia Antipolis, France, <sup>2</sup>IBM Speech Lab, Haifa, Israel, <sup>3</sup>Centre Mémoire de Ressources et de Recherche, CHU de Nice, Nice, France, <sup>4</sup>Association Innovation Alzheimer, Nice, France

**Abstract: Background:** Various types of dementia and Mild Cognitive Impairment (MCI) are manifested as irregularities in human speech and language, which have proven to be strong predictors for the disease presence and progression. Therefore, automatic speech analytics provided by a mobile application may be a useful tool in providing additional indicators for assessment and detection of early stage dementia and MCI.

**Method:** 165 participants (subjects with subjective cognitive impairment (SCI), MCI patients, Alzheimer's disease (AD) and mixed dementia (MD) patients) were recorded with a mobile application while performing several short vocal cognitive tasks during a regular consultation. These tasks included verbal fluency, picture description, counting down and a free speech task. The voice recordings were processed in two steps: in the first step, vocal markers were extracted using speech signal processing techniques; in the second, the vocal markers were tested to assess their 'power' to distinguish between SCI, MCI, AD and MD. The second step included training automatic classifiers for detecting MCI and AD, based on machine learning methods, and testing the detection accuracy.

**Results:** The fluency and free speech tasks obtain the highest accuracy rates of classifying AD vs. MD vs. MCI vs. SCI. Using the data, we demonstrated classification accuracy as follows: SCI vs AD = 92% accuracy; SCI vs. MD = 92% accuracy; SCI vs. MCI = 86% accuracy and MCI vs. AD = 86%.

**Conclusions:** Our results indicate the potential value of vocal analytics and the use of a mobile application for accurate automatic differentiation between SCI, MCI and AD. This tool can provide the clinician with meaningful information for assessment and monitoring of people with MCI and AD based on a non-invasive, simple and low-cost method.

## ARTICLE HISTORY

Received: November 06, 2016  
Revised: July 24, 2017  
Accepted: August 17, 2017

DOI:  
10.2174/1567205014666170829111942

**Keywords:** Alzheimer, dementia, assessment, MCI, speech, audio analysis, machine learning, algorithm.

## 1. INTRODUCTION

As indicated in the World Alzheimer report 2015 [1] there are almost 900 million people aged 60 years and over living worldwide. Rising life expectancy is associated with increased prevalence of chronic diseases like dementia: today 46 million people live with dementia worldwide. It is therefore important to pursue improving awareness and understanding of dementia; providing good quality early diagnosis and intervention for all; improving quality of care from diagnosis to the end of life, using clinical and economic end points [2]. In the field of Alzheimer's disease (AD), criteria have been developed with a better definition of clinical phenotypes and integration of biomarkers into the diagnostic process [3]. A biological marker refers to a broad subcategory of medical signs – that is, objective indications of medical state observed from outside the patient – which can be measured accurately and reproducibly [4].

\*Address correspondence to this author at the CoBTek - Cognition Behaviour Technology - Université de Nice Sophia Antipolis, Centre Mémoire de Ressources et de Recherche- CHU de Nice, Institut Claude Pompidou, 10 rue Molière 06100 Nice, France; E-mail: [alexandra.konig@inria.fr](mailto:alexandra.konig@inria.fr)

This definition underlines the importance of objective measurement tools to assess the health or disease state of an individual. In this area, Information and Communication Technologies (ICT) are of particular interest. Such techniques enable standardized assessments of patients' performance and actions in real time and real life situations [5] but also during clinical research protocols [6]. In fact, it is possible, by using ICT sensors, to obtain a more objective and accurate assessment of behavioral and motor activities of the patient [7-9]. To do such assessment of functions and cognitive performances is one of the main objectives of the European FP7 'Dem@care project' [10].

For the assessment of cognitive and emotional states, speech has been recognized as an important marker till the first psychiatric descriptions [11]. As indicated by Scherer *et al.* [12], speech "is a sensitive output system: slight physiological and cognitive changes potentially can produce noticeable acoustic changes".

It has been demonstrated that various types of dementia significantly affect human speech and language [13, 14, 50].

Therefore, speech can be considered as a source of information for dementia assessment. Dementia affects speech at two levels; the linguistic level ('what is said') and paralinguistic level ('how it is said'). Particularly, in the early phase of dementia, the associated vocal characteristics seem related to temporal parameters of speech, notably longer hesitation times and lower speech rates [15-17].

For this reason, automatic speech processing and machine learning techniques which enable extraction of dementia relevant information from the speech audio signal may be of great interest and has already been used in patients with neurodegenerative disorders or brain injuries [18-20]. Tsanas *et al.*, for instance used speech signal processing algorithms for the detection of Parkinson's disease [21]. Pakhomov *et al.* (2015) investigated the use of automatic speech recognition for longitudinal monitoring of effects of repetitive head trauma on brain function [22]. The automatized analysis of non-linguistic content, such as pause lengths or frequencies, has been demonstrated to provide useful information for cognitive assessment since these markers seem to be sensitive to change with a patient's mental and even emotional state [13, 23-25]. Furthermore, this method of audio capturing is relatively easy and cheap and could allow to carry out assessments and monitoring of disease progress remotely if needed, *e.g.* over the phone.

In a first study we already demonstrated that automatic extraction of vocal markers from speech recordings of elderly and its analysis provided high accuracy rates in classifying up to 87% correctly between healthy subjects (HC) and AD, 80% between mild cognitive impairment (MCI) and AD subjects and 79% between HC and MCI subjects [26]. These results are of great importance since we developed an accurate and cost-effective method supporting clinicians in dementia assessment. However, we aimed to improve these results and overcome emerging shortcomings such as too much background noise or too much variability of the recording quality, by employing an easy to use mobile application which was developed and provided by IBM integrating a short standardized interview protocol that has been designed by the Nice University Memory Center.

The present study aims to demonstrate that automatic speech analysis (ASA) is useful, as an additional objective assessment tool, in clinical practice for diagnosis support purpose. In addition, the study aims to investigate whether there is distinctive pattern of speech related features typical for AD and mixed type dementia. Finally, we aim to define the most sensitive clinical spoken tasks to be used in future work to detect a specific stage and/or type of dementia.

## 2. METHOD

### 2.1. Study Participants

Within the framework of a clinical study carried out for the European research project 'Dem@care', speech recordings were conducted at the Memory Clinic in Nice, France. The Nice Ethics Committee approved the study. Each participant gave informed consent before the assessment. 165 participants aged 65 or older were recruited through the Memory Clinic located at the Institute Claude Pompidou in the Nice University hospital.

All subjects were recorded with a mobile application while performing several short vocal cognitive tasks during a regular consultation. These tasks included verbal fluency, picture description, counting down and free speech tasks.

### 2.2. Clinical Assessments

Each participant underwent a clinical assessment of their cognitive, behavioral and functional status using a battery including: Mini-Mental State Examination (MMSE) [27], the phonemic [28] and semantic verbal fluency [29], the Apathy Inventory [30], the Apathy Diagnostic criteria [31], and the Clinical Dementia Rating Scale (sum of box) [32].

Following the clinical assessment, participants were categorized into four groups: Control participants that complained about having subjective cognitive impairment (SCI) but were diagnosed as cognitively healthy after the clinical consultation, patients with MCI and patients that were diagnosed as suffering from Alzheimer's Disease (AD) and mixed dementia (MD). For the AD group, the diagnosis was determined using the NINCDS-ADRDA criteria [33]. For the related mixed / vascular dementia were diagnosed according to the ICD 10 [34]. For the MCI group, diagnosis was conducted according to Petersen criteria [35]. Participants were excluded if they had any major audition or language problems, history of head trauma, loss of consciousness, psychotic or aberrant motor behavior.

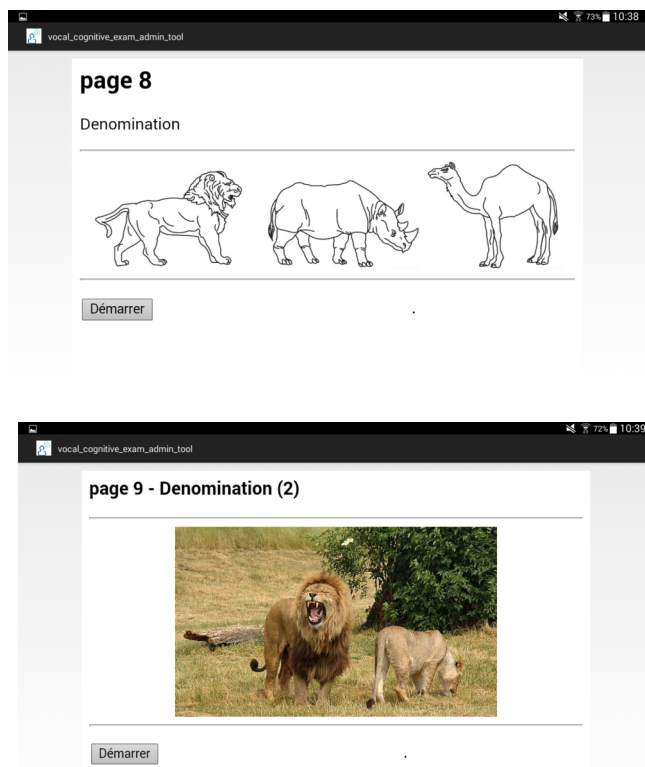
### 2.3. Recording Protocol

Each participant performed a set of six spoken tasks of approx. 15 min in total during a regular consultation with one of the Memory center' clinician who presented the mobile application. The tasks consisted of a sentence-repeating task (two sentences were extracted from the Montreal Cognitive Assessment (MOCA) [36], one from the MMSE [27] and two from our previous study [26]), a denomination task (from the MOCA) [36], two verbal fluency tasks (phonemic and animal naming, which are current tasks in many batteries [23, 28]), a counting backwards task (previous study [26]), and three story-telling tasks (positive, negative and episodic) (see Table 1). These tasks were chosen partly based on the results of our previous study [26] and partly based on the selection of experienced psychologists and speech therapists working at the Memory center. The idea was to use tasks that are already part of common neuropsychological test batteries in order to avoid increasing workload and assessment time for the patient as well as for the clinician.

The vocal tasks were recorded with the 'SmartLav' wearable microphone from the company 'Rode' that was attached to a Samsung Galaxy Tab 3 table on which the developed application was installed.

Each instruction for the vocal task was pre-recorded by one of the psychologist of the center and programmed in the mobile application by IBM. The user interface of the application, designed in collaboration between the clinicians and the engineers from IBM, was kept very simple, only visualizing the instructions, providing images for the denomination task and guiding through the protocol (see Fig. 1). Administration and recording were controlled by the application and facilitated the assessment procedure.

After recording, vocal features were extracted from each spoken task using both the open software tool PRAAT [39] and a set of purposefully-developed signal processing tools.



**Fig. (1).** Screenshot of user interface of mobile application for automatic speech analysis; Picture denomination task (Step 1 above; Step 2 below)

### 2.4. Vocal Features and Statistical Analysis

Demographic variables are described by medians and interquartile ranges. Intergroup comparisons for continuous variables were performed using a nonparametric Kruskal-Willis test given that the distribution of the data was not normal. Categorical testing for gender and education was calculated using the Fischer’s exact test. All statistical analyses of the demographical and neuropsychological data were computed using SPSS 20.0.

The voice recordings are processed in two steps: in the first step, vocal markers are extracted using speech signal processing techniques; a subset of the most relevant vocal markers is then selected based on statistical tests that assess their ‘power’ to distinguish between the different classes (SCI, MCI, AD and Mixed). The second step included training of automatic classifiers for determining the most probable class for which a given set of recorded vocal tasks, representing a person, belong to.

We present results for several different classification scenarios:

- Binary scenarios:
  1. SCI vs. MCI
  2. SCI, vs. AD
  3. SCI vs. Mixed
  4. MCI vs. AD
  5. MCI vs. Mixed
- Multiclass scenarios:
  1. SCI, MCI and AD

**Table 1. List of tasks**

Vocal Task	Instructions	Source
<i>Sentences repetition</i>	I am going to read you a sentence. Repeat it after me, exactly as I say it	2 from MOCA [36] 1 from MMSE [27] 2 from the past collection [26]
<i>Denomination</i> <i>Picture description</i>	Step 1: Tell me the name of this animal (Picture with 3 animals). Step 2: Can you describe me this picture (photography of one of the animal in it natural environment)?	MOCA [36]
<i>Verbal fluency phonemic</i>	Words beginning with the letter F/ In 1 min The voice analysis will only use the first 30s	MOCA [36]. Also exist with other letters in different batteries [23,28]
<i>Verbal fluency semantic</i>	Names of animal / In 1 min The voice analysis will only use the first 30s	Classical task used in different batteries [23, 28]
<i>Counting backward</i>	From 304 to 285 Possible to change for repeated assessment (eg 405, 605)	Classical executive task used in different batteries [37]
<i>Story telling positive</i>	In 1 min can you tell me something about the first pleasant event coming to your mind (if no response prompt with an example)	Adapted from CERAD [38] and IA interview [30]
<i>Story telling negative affect</i>	In 1 min can you tell me something about the first unpleasant event coming to your mind (if no response prompt with an example)	Adapted from CERAD [38] and IA interview [30]
<i>Story telling episodic</i>	Can you tell me what did you do yesterday (or this morning or 2 hours ago....)?	Adapted from CERAD [38] interview

2. SCI, MCI and Mixed
3. SCI, MCI and other (other can be either AD or Mixed).

The vocal markers are calculated in two sub-steps: in the first sub-step, low-level markers (intermediate markers) are calculated directly from the set of recorded vocal tasks. In the second sub-step, the vocal markers for classification are calculated from the low-level markers. There are over 55 different types of vocal markers, calculated separately for each vocal task. As there are 9 different vocal tasks, the total number of vocal markers for a given set of vocal task recordings (from a single person), is about 500. Following selection of the optimal vocal markers, where different vocal markers are selected for each task, only 30 to 50 vocal markers are retained as input for classification. The exact number

of the retained vocal markers depends on the classification scenario.

The following Table 2 lists the types of low-level markers and the associated vocal markers.

The second sub-step of the vocal marker extraction includes testing the “power” of the individual vocal features, and selecting the most representative subset of features for each classification scenario.

Due to the relative small amount of collected data, which is manifested as sparse data in the feature vector space, we used unidimensional feature selection technique. For each classification scenario we choose a subset of the features that in statistical terms best represent this scenario. To explain the feature selection process, we begin with the binary scenarios. For a binary scenario, we choose the subset of fea-

**Table 2. List of types of markers.**

Low Level Marker	Vocal Markers	Description
The durations of silence segments across the audio recording, estimated by speech activity detector	Average, Max, Min, Median, Percentile 15%, Percentile 85%, Sum, Relative part of segments on the first 5 sec, Relative part of segments on the first 10 sec, Number of segments	Higher values designate higher amount of silence periods and lower amount of speech content
The durations of voice segments across the audio recording, estimated by speech activity detector	Average, Max, Min, Median, Percentile 15%, Percentile 85%, Sum, Relative part of segments on the first 5 sec, Relative part of segments on the first 10 sec, Number of segments	Higher values designate higher amount of speech content and lower amount of silence periods
The durations of unvoiced (aperiodic) segments across the audio recording, estimated by pitch detector	Average, Max, Min, Median, Percentile 15%, Percentile 85%, Sum, Relative part of segments on the first 5 sec, Relative part of segments on the first 10 sec, Number of segments	Higher values designate higher amount of unvoiced periods and lower amount of voiced periods
The durations of voiced (periodic) segments across the audio recording, estimated by pitch detector	Average, Max, Min, Median, Percentile 15%, Percentile 85%, Sum, Relative part of segments on the first 5 sec, Relative part of segments on the first 10 sec, Number of segments	Higher values designate higher amount of voiced periods (mostly vowels) and lower amount of unvoiced periods
Time positions of individual words, detected from the signal energy peaks and the pitch envelop (not using verbal content analysis)	Time positions of the first, second,..., the tenth detected words	Lower values designate fast thinking process and rapid speech output ; higher values designate slow thinking process and slower speech output ; relevant mainly to unconnected speech, such as in the Counting and Verbal Fluency vocal tasks
Differences between the time positions of individual words	Average, Max, Min, Median, Percentile 15%, Percentile 85%, relative part of differences below 0.5 sec from all the differences in the first 15 sec, relative part of differences below 1.0 sec from all the differences in the first 15 sec, relative part of differences below 2.0 sec from all the differences in the first 15 sec	Higher relative part of short differences corresponds to presence of clusters, which may hint using associative memory
Time matching curve between the given audio recording of a sentence and a reference audio sentence, calculated by Dynamic Time Warping [as explained in reference...]	Accumulated unmatched segments in the given audio (designate inserted new speech segments), Accumulated unmatched segments in the reference audio (designate missing speech segments), squared error for linear and quadratic approximations	Used for comparing between repeated spoken sentences and the reference spoken sentences Effective for detecting insertions (added parts not present in the reference sentences) and deletions (missing parts from the reference sentences) in the repeated spoken sentences, without using explicit verbal content analysis

tures that have the lowest p-values from a Mann-Whitney test, evaluated on the training data of the classification scenario. The chosen features are the ones with p-values below a predetermined threshold. The specific selected features and their number depend therefore on the specific training data for the classification scenario. Additional details are given in a previous publication from Satt *et al.* (2014) [17]. The cases of multi-class scenarios are handled by the corresponding multiple binary classifiers based on majority vote, and therefore use the binary classifiers and the corresponding feature selection.

The second processing step is the classification: training, testing and validating the classification accuracy. The binary classification is handled as follows: we use Support Vector Machine classifiers. For validation we used repeated sub-sampling, which is suitable for the relatively small amount of collected data. The cases of three-class scenarios are implemented by the corresponding multiple binary classifiers and majority decision rule.

### 3. RESULTS

#### 3.1. Characteristics of the Participants

Since the distribution of data was nonparametric, results are reported in medians and interquartile ranges. Table 3 shows the clinical and demographic data of the participants.

A Kruskal-Wallis test showed that there was a statistically significant difference in age between the groups,  $\chi^2(3) = 23.312, p = .000$ , in the MMSE score,  $\chi^2(3) = 120.699, p = .000$ , in the verbal fluency (f),  $\chi^2(3) = 36.637, p = .000$ , with a mean score for SCI = 14.7, MCI = 11.1, AD = 7.1 and mixed dementia = 12.2, in the verbal fluency (animals),  $\chi^2(3) = 60.971, p = .000$ , with a mean score for SCI = 19.3, MCI = 14.1, AD = 7.1 and mixed dementia = 9.2, in the Apathy Inventory,  $\chi^2(3) = 68.811, p = .000$  with a mean score for SCI = 0.3, MCI = 1.3, AD = 4.1 and mixed dementia = 4.9, and in the Clinical Dementia Rating Scale (sum of boxes),  $\chi^2(3) = 106.373, p = .000$  with a mean score for SCI = 0.3, MCI = 1.8, AD = 7.8 and mixed dementia = 7.9. The categorical testing with the Fischer's exact test showed

**Table 3. Characteristics and group comparisons for SCI, MCI, AD and mixed subjects. Group comparisons were made using Kruskal-Wallis test.**

	All subjects				AD				Mixed				MCI				SCI			
	mean	median	Q1	Q3	mean	median	Q1	Q3	mean	median	Q1	Q3	mean	median	Q1	Q3	mean	median	Q1	Q3
<b>N</b>		165				27				38				44				56		
<b>Gender</b>																				
<b>Female</b>		108				15				17				25				51		
<b>Male</b>		57				12				21				19				5		
<b>Age (y)</b>	76.2	77*	72	81	79.1	80	76	84	79.1	80	76	84	76.3	77	72	81	72.3	72	70	77
<b>Education</b>																				
<b>NA</b>		12				0				0				0				12		
<b>None</b>		3				3				0				0				0		
<b>Primary</b>		39				5				18				10				6		
<b>Secondary</b>		69				14				13				25				17		
<b>Superior</b>		42				5				7				9				21		
<b>MMSE</b>	23.9	26*	20	28	18.2	19	14	21	18.6	19	16	22	26.1	27	25	28	28.7	29	28	30
<b>Fluency F</b>	11.4	10*	6	14	7.1	7	4	10	12.2	7	3	10	11.1	12	7	15	14.7	9	0	15
<b>Fluency animals</b>	12.5	12*	8	17	7.1	7	4	10	9.2	10	5	12	14.1	13	9	18	19.3	11	0	19
<b>AI</b>	2.6	1.5*	0	5	4.1	4	1	6	4.9	5	3	6	1.3	0	0	2	0.3	0	0	0
<b>Apathy criteria (yes)</b>		25%				48%				61%				11%				0%		
<b>CDR SOB</b>	4.1	2.5*	1	6	7.8	6	4	9	7.9	7	5	10	1.8	2	1	2	0.3	0	0	1

**Abbreviations:** SCI: subjective cognitive impairment, MCI: mild cognitive impairment, AD: Alzheimer's Disease, Mixed: mixed type dementia, MMSE: Mini Mental State Examination, AI: Apathy Inventory, CDR SOB: Clinical dementia rating-Sum of Boxes.

\* $P < .000$

significant difference for gender with  $\chi^2(3) = 26.087$ ,  $p = .000$ , for education levels with  $\chi^2(12) = 62.191$ ,  $p = .000$ , and for apathy  $\chi^2(9) = 92.065$ ,  $p = .000$ .

### 3.2. Classification and Analysis

We evaluated the classification accuracy for several classification scenarios as follows:

- Binary scenarios
  1. SCI vs. AD 92%
  2. SCI vs. Mixed 92%
  3. SCI vs. MCI 86%
  4. MCI vs. AD 86%
  5. MCI vs. Mixed 82%
- Multiclass scenarios
  1. SCI, MCI and AD 78%
  2. SCI, MCI and Mixed 74%
  3. SCI, MCI and

other (other can be either AD or Mixed)

The classification accuracies above correspond to the points along the ROC curves where the specificity is equal to the sensitivity. For different uses and purposes, the classifiers can be tuned to support different tradeoffs of specificity vs. sensitivity. Detailed explanation is given in our previous publication [17].

The following table summarizes the vocal marker selection results (Table 4), by vocal task and by classification scenario. The total number of selected vocal markers is between 30-50, depending on the classification scenario. Larger numbers of selected features suggest higher importance of the specific vocal task to the particular classification scenario.

Table 4. demonstrates the relative importance of the different vocal tasks, for supporting the different classification scenarios:

- The countdown task is in particular powerful for separating between SCI and other classes, although it helps also for distinguishing among other classes. See Fig. (2) for a visualization example of the countdown task performance of and Alzheimer patient vs. a subjective cognitive impairment subject.
- The fluency tasks are in particular powerful for separating the dementia cases, both AD and Mixed/vascular, from MCI.
- The various continuous speech tasks collectively, are important as well. They include the different stories, and the picture description. While each one may not be that powerful, they contribute collectively to a high number of powerful vocal markers, with different combinations across different classification scenarios.
- The short task of naming animals has no significant importance. It is our observation that short tasks are prone to high statistical errors, which makes them less powerful for classification.

- Finally, the sentence repeating task does not contribute at all to the classification accuracy. We observed that SCI person and patients alike have no real difficulty repeating the short sentences we used in a good manner. We suggest a hypothesis that longer sentences can be found helpful in classification scenarios, although we haven't tested it.

It should be noted that we have tested additional types of vocal markers, such as voice quality based on harmonic to noise ratio. We have not found a sufficiently significant correlation between these additional features and the classes (SCI, MCI, AD and Mixed). Speech-rate type of markers is modeled implicitly by the vocal markers we already used.

## 4. DISCUSSIONS

Decline in cognitive functioning affects speech production in different ways. Our performed analysis demonstrated the potential value of vocal cognitive tasks recorded and analyzed by a mobile application for accurate automatic differentiation between SCI, MCI, AD and Mixed dementia. Our highly accurate classification results can provide evidence that it is feasible to use this technique in a clinic setting, which is indeed non-invasive, simple and low-cost.

The present results are also in line with the concept indicating that different stages of dementia exhibit specific patterns of linguistic difficulties [40]. This seems confirmed by our results and the more cognitive decline is advanced, the less cognitive effort is required in a vocal task to extract significant features, thus the more salient the differences in vocal expression. However, main interest is placed currently on the detection of early markers of cognitive decline.

In this field there are studies investigating the concerning neuropathological brain changes. Results indicated that areas implicated in verbal working memory performance and language processing were affected, notably disconnection (impairments of structural fiber tract integrity) of prefrontal to posterior brain regions through the posterior corpus callosum [41-43]. It seems of interest to develop research in order to look at the relation between anatomical parameters and speech production analysis techniques.

Currently, the inadequacy of existing methods combined with biased evaluations, points to a need for objective and systematic assessment tools and researchers aim to provide novel solutions. Clinical expertise and literature review indicates that ICT are not yet able to provide a direct diagnosis of AD and related disorders, but can supply additional information for the assessment of specific domains (behavior, cognition, activity of daily living) [5]. This information can contribute with other clinical and biological data to earlier diagnosis of AD and related disorders. Several studies using ICT in the assessment of different domains show potential benefits of using ICT in clinical practice. As our study demonstrates, automatic audio analysis could help to earlier identify individuals that are more likely to develop dementia, which may allow clinicians to provide earlier timely care, treatment (pharmacological as non-pharmacological) and support, which will in turn reduce health care costs [6].

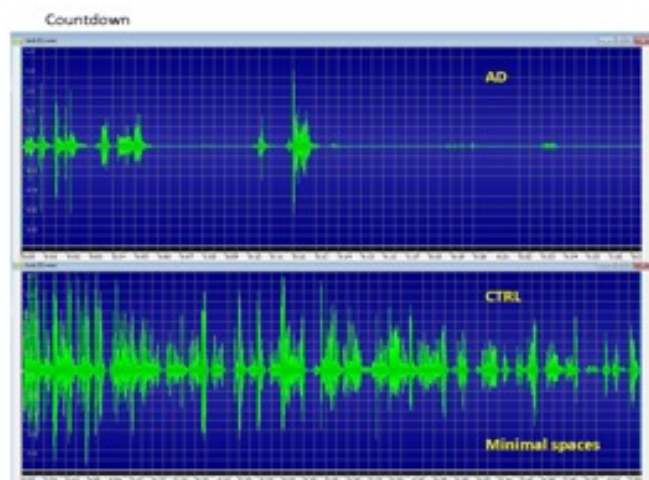
Namely, drug research focuses at the moment on targeting patients at the very early stages of the disease when

**Table 4. Vocal marker selection results.**

Vocal Task	Key features – examples as selected by threshold on p-values	Classification scenarios				
		Relative amount of selected features, as a measure for the importance of the vocal task for each classification scenario				
		HC vs. MCI	HC vs. AD	HC vs. Mixed	MCI vs. AD	MCI vs. Mixed
Animal fluency	Time locations of first words Distribution of time between words	0	2	3	8	9
Pictures of animals	Time locations of first words Distribution of time between words	0	1	1	1	0
Picture description	Durations of silence segments Durations of voice segments Durations of unvoiced segments Durations of voiced segments	3	1	4	4	2
F fluency	Time locations of first words Distribution of time between words	0	0	6	6	4
Countdown	Durations of silence segments Durations of voice segments Durations of unvoiced segments Durations of voiced segments	16	22	11	5	7
Positive story	Durations of silence segments Durations of voice segments Durations of unvoiced segments Durations of voiced segments	3	2	0	2	3
Negative story	Durations of silence segments Durations of voice segments Durations of unvoiced segments Durations of voiced segments	4	3	4	1	3
Episodic story	Durations of silence segments Durations of voice segments Durations of unvoiced segments Durations of voiced segments	6	7	3	2	3
Repeating sentences	The best vocal markers in terms of low p-values are the estimations of insertions (including pauses) and deletions; higher values designate poor sentence repeating ; however, no vocal marker from this task was ever selected	0	0	0	0	0

memory functions are still preserved. This means that the use of ICT could have a direct beneficial effect on the selection of people for the enrolment in clinical trials in the broader

population, leading ultimately to a reduction of the total burden for society.



**Fig. (2).** Visualization of countdown task performance – Alzheimer patient vs. a subjective cognitive impairment subject

Our work relates to similar designed studies like the one from Fraser *et al.* [44] who obtained classification accuracies of over 81% in distinguishing individuals with AD from those without. Their results are based on short language samples of a picture description task. Nonetheless, they only opted for differentiating between AD and non AD patients, whereas our aim was particularly to include MCI patients in the analysis in order to find vocal markers present already in the early pre-dementia stage. However, we can conclude that current machine learning and automatic speech analysis may become increasingly useful in assessment of patients across the AD spectrum which seems supported by other studies [48-49].

Ultimately, similarly to Yu *et al.* (2015) [45], we are aiming for an approach for remote cognitive monitoring of elderly even over the phone in order to allow general regularly frontline screening in the broad population. This would help to identify early people that may be of risk to convert to dementia and therefore require early intervention provided at stages when it may be more effective. The strength in the envisioned approach relies on the fact that only non linguistic features are taken into consideration in the analysis and thus, allow a wide international application independently of the spoken language since we aimed for language-independent features only.

The early detection challenge needs to carefully choose the technical aspect such as the automatic extraction of the vocal features [46] but also the clinical characteristics of the task to be used. If we aim to detect early signs of cognitive decline in the voice, it is of great importance to choose a task that requires sufficient cognitive effort in order to obtain significant differences in the speech features. If we look only into natural speech characteristics, research so far has been mostly able to achieve high classification accuracy between dementia and non-dementia subjects [18, 44]. Certain studies reveal that dementia patients show increased amount of pauses in speech which may be attributable to retrieval difficulties from the lexical-semantic stock but may reflect as well other cognitive processing deficits since it increases with cognitive effort [15, 47, 17]. Pistono *et al.* (2015) [47] analyzed pausing behavior (frequency, duration, and location) in MCI patients during an

ecological episodic memory task with the results that they did not produce more pauses than controls but made more in between-utterance pauses maybe representing processing speed for memory retrieval. However, results obtained in the present study are in line with the idea to include vocal tasks with a minimum of cognitive effort in study protocols that aim to assess speech parameters particularly in an early dementia population.

A major limitation of the study is the important heterogeneity of our population groups, as the statistical tests showed significant difference in age between the groups, as well as gender and education. This is due to the fact that the recruited population were consulted at the memory clinic, in thus in itself represents a broad group of people coming from different backgrounds and education levels. It should be further noted that we did not recruit healthy participants from the general elderly population, but rather we limited the group to include persons that came in for clinical consultation and had subjective complaints. However, it reflects the expected scenario that our technology is likely to serve: people already suffering from some (subjective) level of cognitive or functional problem, though below the level of clinical MCI.

Future research should focus on carrying out longitudinally recording of speech data of patients over time for the purposes of early dementia detection as well as on performing a larger international cross-linguistic study in order to define language-independent features that seem to indicate over time the risk to decline cognitively.

#### **ETHICS APPROVAL AND CONSENT TO PARTICIPATE**

Not applicable.

#### **HUMAN AND ANIMAL RIGHTS**

No Animals/Humans were used for studies that are base of this research.

#### **CONSENT FOR PUBLICATION**

Not applicable.

#### **CONFLICT OF INTEREST**

Authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

#### **ACKNOWLEDGEMENTS**

This study was supported by grants from the FP7 Dem@care project, by the Innovation Alzheimer association, by the IBM Speech Lab in Haifa, Israel. The CoBTeK (Cognition – Behaviour – Technology) Research Unit from the Nice Sophia-Antipolis University (UNS) located at the Institut Claude Pompidou in the Edmond and Lilly Safra research center, the Nice CHU Memory Center (platform of the CIUS). Thank to the foundation Claude Pompidou for providing such facilities. Final great thanks to all participants of the study.



## REFERENCES

- [1] World Alzheimer Report 2015; The Global Impact of Dementia: An Analysis of prevalence, incidence cost and trends published by Alzheimer's Disease International (ADI), London. September 2015. Copyright © Alzheimer's Disease International.
- [2] Kenigsberg PA, Aquino JP, Berard A, Gzil F, Andrieu S, Banerjee S, *et al.* Dementia beyond 2025: Knowledge and uncertainties. *Dementia (London)* 15(1): 6-21 (2016).
- [3] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, *et al.* Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol* 9(11): 1118-1127 (2010).
- [4] Strimbu, K, Tavel, JA. What are biomarkers? *Curr Opin HIV AIDS* 5(6): 463-466 (2010).
- [5] Robert PH, König A, Andrieu S, Bremond F, Chemin I, Chung PC, *et al.* Recommendations for ICT use in Alzheimer's disease assessment: monaco CTAD expert meeting. *J Nut Health Aging* 17(8): 653-660 (2013).
- [6] König A, Sacco G, Bensadoun G, Bremond F, David R, Verhey F, *et al.* The role of information and communication technologies in clinical trials with patients with Alzheimer's disease and related disorders. *Front. Aging Neurosci* 7: 110 (2015).
- [7] Yakhia M, König A, van der Flier WM, Friedman L, Robert P. H, *et al.* Actigraphic motor activity in mild cognitive impairment patients carrying out short functional activity tasks: comparison between mild cognitive impairment with and without depressive symptoms. *J Alzheimers Dis* 40: 869-875 (2014)..
- [8] Nagels G, Engelborghs S, Vloeberghs E, Van Dam D, Pickut, BA, De Deyn PP. Actigraphic measurement of agitated behaviour in dementia. *Int J Geriatr Psychiatry* 21 : 388-393 (2006)..
- [9] David R, Mulin E, Friedman L, Le Duff F, Cygankiewicz E, Deschaux O, *et al.* Decreased daytime motor activity associated with apathy in Alzheimer disease: an actigraphic study. *Am J Geriatr Psychiatry* 20: 806-814 (2012)..
- [10] <http://www.demcare.eu/>
- [11] Kraepelin E. (1921). *Manic-depressive Insanity and Paranoia*. New York: Arno Press, 1976.
- [12] Scherer KR. Vocal affect expression: A review and a model for future research. *Psychol Bulletin* 99: 143-165 (1986)..
- [13] Appell J, Kertesz A, Fisman M. A study of language functioning in Alzheimer patients. *Brain Lang* 17: 73-91 (1982)
- [14] Reilly J, Rodriguez AD, Lamy M, Neils-Strunjas J. Cognition, language and clinical features of non-Alzheimer's dementias: an overview. *J Commun Disord* 43(5): 438-452 (2010).
- [15] Hoffmann I, Németh D, Dye C, Pákási M, Irinyi T, Kálmán J. Temporal features of spontaneous speech in Alzheimer's disease. *Int J Speech Lang Pathol* 12: 29-34 (2010).
- [16] 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* 19(7): 2081-2090 (2011).
- [17] Satt A, Hoory, R, König, A, Aalten, A, Robert, PH. Speech-based automatic and robust detection of very early dementia. In *Proceedings of the INTERSPEECH Conference 2014*, September 14-18, 2014, Singapore
- [18] Bucks RS, Singh S, Cuerden JM, Wilcock GK. Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analysing lexical performance. *Aphasiology* 14(1): 71-91 (2000).
- [19] Canning SJ, Leach L, Stuss D, Ngo L, Black, SE. Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. *Neurology* 62(4): 556-62 (2004).
- [20] Forbes-McKay KE, Venneri A. Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurol Sci* 26(4): 243-54 (2005).
- [21] Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Transactions Bio-med Engineering* 59: 1264-1271 (2012).
- [22] Pakhomov SV, Marino SE, Banks S, Bernick C. Using Automatic Speech Recognition to Assess Spoken Responses to Cognitive Tests of Semantic Verbal Fluency. *Speech Commun* 75: 14-26 (2015).
- [23] Yu B, Quatieri TF, Williamson JR, Mundt, JC. (2014). Prediction of cognitive performance in an animal fluency task based on rate and articulatory markers. In *Fifteenth Annual Conference of the International Speech Communication Association*.
- [24] Roark B, Mitchell M, Hosom JP, Hollingshead K, Kaye J. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Trans Audio, Speech, Lang Proc* 19(7): 19(7): 2081-2090 (2011).
- [25] Trevino A, Quatieri TF, Malyska N. Phonologically-based biomarkers for major depressive disorder. *EURASIP Journal on Advances in Signal Processing: Special Issue on Emotion and Mental State Recognition from Speech* 42: 2011-2042, (2011).
- [26] König A, Satt A, Sorin A, Hoory R, Derreuxaux A, Manera V, *et al.* Automatic speech analysis for the assessment of pre-demented and Alzheimer patients. *Alzheimer's Dementia: Diagnosis, Assessment & Disease Monitoring* 1: 112-124 (2015b).
- [27] Folstein M, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12: 189-198 (1975).
- [28] Loonstra AS, Tarlow AR, Sellers AH. COWAT metanorms across age, education, and gender. *Appl Neuropsychol* 8(3): 161-6 (2001).
- [29] Ardila A, Ostrosky-solis F, Bernal B. Cognitive testing toward the future: the example of semantic verbal fluency (ANIMALS). *Intern J Psychol* 41 (5): 324-332 (2006).
- [30] Robert PH, Claret S, Benoit M, Koutaich J, Bertogliati C, Tible O, *et al.* The apathy inventory: assessment of apathy and awareness in Alzheimer's disease, Parkinson's disease and mild cognitive impairment. *Int J Geriatr Psychiatry* 17(12): 1099-105 (2002).
- [31] Mulin E, Leone E, Dujardin K, Delliaux M, Leentjens A, Nobili F, *et al.* Diagnostic criteria for apathy in clinical practice. *Int J Geriatr Psychiatry* 26(2): 158-65 (2011).
- [32] O'Bryant SE, Waring SC, Cullum CM, Hall J, Lacritz L, Massman PJ, *et al.* Staging dementia using clinical dementia rating scale sum of boxes scores: a texas alzheimer's research consortium study. *Arch Neurol* 65(8): 1091-1095 (2008).
- [33] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34(7): 939-44 (1984).
- [34] World Health Organisation: ICD-10 Classifications of Mental and Behavioral Disorder: Clinical Descriptions and Diagnostic Guidelines. Geneva, Switzerland, World Health Organisation, 1992
- [35] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 56(3): 303-8 (1999).
- [36] Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, *et al.* The montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53: 695-699 (2005).
- [37] Reisberg B, Ferris SH. Brief cognitive rating scale. *Psychopharmacol Bulletin* 24(4), 629-636 (1988).
- [38] Tariot PN, Mack JL, Patterson MB, *et al.* The CERAD Behavioral Rating Scale for Dementia. *Am J Psychiatry* 152: 1349-1357 (1995).
- [39] <http://www.fon.hum.uva.nl/praat/>
- [40] 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* 7: 195 (2015).
- [41] Teipel S, Ehlers I, Erbe A, Holzmann C, Lau E, Hauenstein K, *et al.* Structural connectivity changes underlying altered working memory networks in mild cognitive impairment: a three-way image fusion analysis. *J Neuroimaging* 10: 101 (2014).
- [42] Zhang Y, Schuff N, Camacho M, Chao LL, Fletcher TP, Yaffe K, *et al.* MRI markers for mild cognitive impairment: comparisons between white matter integrity and gray matter volume measurements. *PLoS ONE* 8: e66367 (2013)..
- [43] Liu Y, Spulber G, Lehtimäki KK, Könönen M, Hallikainen I, Gröhn H, *et al.* Diffusion tensor imaging and tract-based spatial statistics in Alzheimer's disease and mild cognitive impairment. *Neurobiol Aging* 32: 1558-1571 (2011)..
- [44] Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimers Dis* 49(2): 407-22 (2015).
- [45] Yu B, Quatieri TF, Williamson JR, Mundt JC. Cognitive impairment prediction in the elderly based on vocal biomarkers. *Proceed-*

- ings of Interspeech 2015. September 6-10, 2015, Dresden, Germany.
- [46] 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, in Interspeech 2015 Tutorials & Main Conference, Germany: Dresden (2015).
- [47] Pistono A, Jucla M, Barbeau EJ, Saint-Aubert L, Lemesle B, Calvet B, *et al.* Pauses during autobiographical discourse reflect episodic memory processes in early Alzheimer's disease. *J Alzheimer's Dis* 50(3) (2015).
- [48] López-de-Ipiña K, Martínez-de-Lizarduy U, Calvo PM, Mekyskac J, Beitiad B, Barroso N, *et al.* Advances on automatic speech analysis for early detection of Alzheimer Disease: a non-linear multi-task approach. *Curr Alzheimer Res* (2017).
- [49] Tóth L, Hoffmann I, Gosztolyac G, Vinczec V, Szatlóczkid G, Bánrétib Z, *et al.* A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Curr Alzheimer Res* (2017).
- [50] Misiewicz S, Brickman AM, Tosto G. Prosodic impairment in dementia: review of the literature. *Curr Alzheimer Res* (2017).

DISCLAIMER: The above article has been published in Epub (ahead of print) on the basis of the materials provided by the author. The Editorial Department reserves the right to make minor modifications for further improvement of the manuscript.

**PMID: 28847279**