Automatic tool for screening of cognitive impairments

Cristiana Filipa Lopes Amorim

IST – Instituto Superior Técnico
L2F – Spoken Language Systems Laboratory – INESC ID Lisboa
Rua Alves Redol 9, 1000-029 Lisboa, Portugal
cristiana.f.l.amorim@gmail.com

Abstract. The progression of Alzheimer’s Disease (AD), which is the most common form of dementia, can be slowed down if detected early. Several screening tests have been developed for this purpose, Mini-Mental State Examination (MMSE) being one of the most widely used. However, such tests need to be administrated by a specialized therapist and typically consist of frequent meetings between patients and therapists, which difficulties their applicability. This thesis aims to develop an automatic web-based tool that enables patients to perform screening tests at home without the assistance of therapists and, simultaneously, help therapists in diagnosing and monitoring the disease. The tool was implemented by adapting an existing platform for aphasia treatment, known as Virtual Therapist for Aphasia Treatment (VITHEA). This platform integrates the L2F in-house Speech and Language Technologies (SLT). Overall, the obtained results were satisfactory, which suggest that the platform may be useful and reliable as a screening tool.

1 Introduction

The Alzheimer’s Disease (AD) is a neurodegenerative disease which represents 60 to 70% of the dementia cases in Portugal [6,5,8]. However, its first signs can go unnoticed [6,2,5,3].

Typically, AD is known to cause alterations of memory and of spacial and temporal orientation [4,2,8]. Furthermore, AD increases dramatically with age and it has no cure. Nevertheless, an early diagnosis may slow down its progression by enabling a more effective treatment [3]. For this purpose, several neuropsychological tests exist on the literature.

The traditional assessment raises, however, some issues. First of all, a therapist must be present to apply the neuropsychological test to the patient. Second, it is important that the assessment is continuous, i.e., the tests should be applied over time and not just once. This implies that, each time the patient needs to be evaluated by the therapist, the patient has to go to a specialized place (such as a hospital or a clinic) or, alternatively, the therapist has to go to the patient’s home, which is very costly and inconvenient.
This Master Thesis aims to select a set of neuropsychological tests and adapt them to an automatic web-based system that integrates Speech and Language Technologies (SLT), which will serve to therapists as a tool to assist in the diagnosis and monitorization of cognitive impairments, with a main focus on AD. With this tool, the patient will be allowed to perform the tests at his/her home without the assistance of a therapist. Furthermore, SLT will enable a natural interaction between the patient and the system, since in this way the patient can speak directly to the system and this last can provide feedback, simulating a real therapist. This kind of interface is appropriate for people with AD that may have some motor limitations or orientation problems due to the disease (which may make it difficult for them to use devices like mouse and keyboard). The tool will be implemented for the Portuguese population.

The framework for developing this tool is VITHEA, an on-line platform used for aphasia treatment that incorporates SLT to provide word naming exercises. For this to be possible, the system resorts to a keyword spotting technique which consists of detecting a certain set of words by using a competing background model with the keywords model. This platform is used daily by patients and speech therapists and has received several awards from both the speech and the health-care communities. The great success achieved with the VITHEA platform motivated its use as a baseline for the present thesis. Furthermore, we thought to be of great interest to have a single platform that includes exercises for several pathologies.

Due to space constraints, in this document we focus on the developed work for this thesis. For a detailed description of the state of the art and related work, please refer to sections 2 and 3.1 of the MsC thesis document [1]. In the following, Section 2 reports on the neuropsychological tests that where selected for further implementation. Section 3 details the main modifications required to adapt the VITHEA platform to the selected neuropsychological tests. In Section 4, the focus is on briefly describing how each type of exercise was concretely implemented. Then, Section 5 introduces an overview of the evaluation process. Finally, Section 6 presents the conclusions and future work.

2 Selected Neuropsychological Tests

For the present thesis, we chose to implement a set of the neuropsychological tests that were presented as Related Work for this thesis. This choice fell upon the most commonly used tests for the Portuguese population, concretely the Mini-Mental State Examination (MMSE) and the Alzheimer’s Disease Assessment Scale - Cognitive Subscale (ADAS-cog). These tests are also interesting from the point of view of the diversity of questions. However, since the focus of this work is related to SLT, we decided to implement only the questions that comprise a speech component, thus resulting in a partial implementation of the selected neuropsychological tests. Anyway, some of those questions would be impossible to implement in its original form. Particularly, MMSE and ADAS-cog comprise exercises of several types: orientation to time, place and person; evo-
cation; attention and calculation; naming objects and fingers; repeat sentences; and word recognition.

Another test that is commonly used by therapists is the Animal Naming test. This is a test of the evocation type which consists of naming animals in a one-minute interval. Besides, it increases the sensitivity of the MMSE and, thus, we also chose to implement it for the present thesis. Moreover, this is an interesting test as regards the automatic speech evaluation.

3 Main Modifications

In order to support the new tests, the VITHEA platform had to be adapted.

First of all, a cognitively impaired person may present some level of disorientation. Therefore, an interface which minimizes the usage of mouse and the number of clicks would be desirable. In this context, we decided to automatize the recording start. Our approach consisted of starting the recording at the beginning, as soon as the virtual therapist starts asking the question. This implied an imposition of a timestamp to mark the moment when the virtual therapist finishes speaking, which is used afterwards to trim the recorded audio in the right place.

For the exercises used in aphasia treatment and already present in VITHEA, a feedback is provided for each stimuli shown, either positive or negative. Contrarily, for neuropsychological tests only a neutral feedback should be provided during the test application and the global score is calculated at the end of the test. This led to a modification for this kind of exercises, in which the positive or negative feedback is substituted by a neutral feedback. At each point, the feedback choice is done randomly.

The automatic recording and the neutral feedback motivated 2 more changes. Since the neuropsychological tests only allow one attempt for each question, and since the score is displayed only at the end, it makes no sense that the patient can listen to the answer he/she gave before completing the test. Therefore, the playback button was disabled for this kind of tests. Besides, as a means of reducing even more the usage of the mouse and the number of clicks and to make the progression of questions more fluid, we decided to implement the automatic advancement of questions for the neuropsychological test. In this way, as soon as the user stops the recording and the submission of his/her answer has finished, the system provides the neutral feedback, if applicable, and then automatically advances to the next question. These alterations contributed to a simplified interface, which is suitable for aged people, specially if cognitively impaired.

The previous changes raised another question: providing always a neutral feedback for all questions may become boring, frustrating and useless for the patient. Besides, for a test comprising a diversity in terms of questions this would become even odd and slow for similar consecutive questions and accentuates the lack of cohesion between them. Since the questions are organized in themes, the solution was to provide a neutral feedback only when the theme changes and not provide a feedback at all in the other cases.
Another implemented modification was the slowing of the speech rate of
the virtual therapist. This was motivated not only by the condition of cognitive
impairment, but also by the ageing factor which often results in a diminished au-
ditory condition. In this way, we tried to make the speech more understandable,
but without compromising its naturalness and without giving the perception of a
sad voice (often associated with a slower speech). After some experimentations,
we found the optimal speech rate that complies with these requirements.

The inclusion of new exercises implies the addition of new categories in the
database. Moreover, it was necessary to create a new home page for the patient
application module in order to distinguish between the exercises for aphasia
treatment and the neuropsychological tests.

Regarding the clinician application module, the form to create and edit the
exercises had to be changed to support the new tests. Due to the diversity
of questions, and contrarily to the exercises for aphasia treatment, the same
test must support several types of stimuli and not just once. Since the type
of stimuli depends on the question’s theme, the form was adapted to change
accordingly to it by showing the relevant fields and hiding the remaining ones
as a means of avoiding errors. Furthermore, a new field was added to support an
optional instruction for the stimuli. This may be useful to explain a more complex
exercise before presenting the stimuli and asking the question associated to it.
To support this, the patient application module had to be adapted to enable the
virtual therapist to speak the instruction. Basically, if provided, the instruction
is spoken by the virtual therapist while the stimuli is hidden. As soon as the
instruction is given, the stimuli is presented and the virtual therapist asks the
question associated to it.

Additionally, the form to create and edit the patient’s profile was also modi-
fied to include information that is required for implementing the tests (e.g.,
county and town) and to contain information related to AD.

Finally, during the application of a neuropsychological test, the scores are
individually calculated for each question. Concretely, after the answer has been
processed by the Automatic Speech Recognition (ASR) system, the platform
computes the maximum score allowed for the current question, as well as the
score obtained by the patient in the question. These results are stored in the
database. At the end of the test, both the maximum scores and the obtained
scores for each question are summed separately to obtain a global score in the
form \( \text{score} / \text{maxScore} \) (e.g., a score of 18/22). This result can then be consulted
by the patient. In order to follow the patient’s progress, each time an evaluation
test is performed the platform sends an e-mail with a summary of the patient’s
performance to the therapist assigned to him/her.

4 Automated exercises

Since the selected neuropsychological tests comprise common or similar exercises,
we may approach its concrete implementation organized by type of exercise
rather than per test. It is worth mentioning that each type of exercise has set
different challenges and thus different solutions were required to address these challenges. Furthermore, after a first meeting with the therapists, the interface has changed and so only the final version of the implemented exercises would be approached here. For detailed information please refer to the Msc document [1].

4.1 Repetition

The repetition exercise consists of repeating a sentence. This exercise could be easily implemented as a keyword spotting exercise, just like the ones for aphasia treatment. Thus, it was only necessary to create a stimuli of the auditory type and add the sentence as the correct answer associated to it. The maximum score for this exercise is 1, which corresponds to a sentence correctly repeated in its entirety.

4.2 Attention and calculation

The idea of this exercise is to successively subtract 3 beginning on 30 until 5 answers are given. The result was a set of 5 different stimuli, each one asking separately for a specific calculation. In terms of score, this exercise was implemented as a normal keyword spotting exercise, with a maximum score of 1 (for each stimuli) corresponding to a correct answer.

4.3 Naming objects and fingers

This exercise is similar to the ones already existing in the VIThea platform, with the major difference that now some of the questions allow an optional semantic cue. This was implemented by adding a timer in the component responsible for the answer’s recording and by making the virtual therapist to spoken the cue after 20 seconds if no answer has been sent yet. Additionally, the timestamp used to trim the audio needs also to be updated.

4.4 Orientation to time, place and person

These kind of exercises comprise questions such as the year, day, month, contry, town you live, hour, among others. These are dynamic questions in the sense that there is no universal answer to each question as this changes depending on the time, place and person. The solution was to provide several pre-compiled language models that were carefully structured so that, at any time, the platform knows which is the right model to chose. For the questions of orientation to person, the necessary information is acquired at the time of the creation of the user profile and then it is used to automatically generate the corresponding language models.

The majority of the questions was implemented as normal keyword spotting exercises. However, for the day and hour, it was necessary to pre-compile an unique language model for each of the questions. These language models are
rule-based and use tags as a means to convert the words into numbers, which is necessary to allow the evaluation of the answer’s correctness by comparison with the system date or time. Concretely, if a set of words is recognized the corresponding numeric tag will be returned, e.g. day am (one) is associated with tag 1.

4.5 Word recognition

The word recognition exercise consists of presenting the patient a list with 12 words to learn, one at a time and written in block letters on white cards. The learning process is made by asking the patient to read each word aloud and try to remember it. Then, a new list with 24 words is shown in the same way. This new list contains the 12 original words of the learning list, plus 12 new distracting words that are carefully chosen in terms of phonetic similarity and semantic meaning. For each word, the patient is then asked to indicate whether it was on the learning list or not. This whole process is repeated in 3 trials.

The form that is used to create or edit the stimuli was modified specifically for this exercise to enable the creation or edition of a complete trial based solely on text. This means that the therapist must indicate the learning list and the list of words to recognize and then the platform automatically generates all the textual stimuli as well as the corresponding correct answer.

Just like the day and hour questions, the word recognition exercise requires a rule-based language model with tags. The idea is to recognize the most commonly used answers, even positive, negative or neutral, and then compare the recognized answer with the correct answer which was automatically generated upon the stimuli creation. In this context, the tags play an essential role by standardizing the recognized answer returned by the ASR system, presenting it in the form of a sim (yes) or náo (no) answer, or discarding it if the patient says he/she does not know the answer.

4.6 Evocation

Generally speaking, an evocation exercise consists of recalling a series of words, whether they have been previously learned or if they are subject to compliance with certain requirements. In either cases, the spoken answers produced for this kind of exercises are commonly followed by filled pauses, i.e. hesitation sounds. For this reason, in order to implement these exercises we adopted a keyword spotting approach, but incorporating a model to represent the filled pauses. In terms of score, the calculation is processed by considering the number of keywords correctly produced, without repetitions.

For MMSE, the evocation exercise consists of the immediate and delayed recall of 3 words. This was implemented with an auditory stimuli for the immediate recall task and with a textual stimuli for the delayed recall task. However, a new field is required to support the insertion of the list of correct answers.

The presentation of the evocation exercise that belongs to the ADAS-cog is very similar to the word recognition exercise. Basically, this exercise consists of
the immediate recall of a list with 10 words that were previously learned in the same manner as the word recognition exercise, and the whole process is repeated in 3 trials. Thus, just like the word recognition exercise, the form that is used to create or edit the stimuli was also modified specifically for this exercise to allow the automatic generation of the stimuli for a complete trial. Therefore, the therapist must indicate the list of words to learn and recall, which also corresponds to the list of correct answers for the immediate recall task itself.

Lastly, the Animal Naming exercise is the most challenging one. First of all, we expect a higher number of hesitations, filled pauses and repetitions. On the other hand, and contrarily to the other language models which are based on a limited domain with a reduced set of words, the Animal Naming exercise comprises a more extended domain. Particularly, the keywords list represents the names of admissible animals that will be accepted by the ASR system. Theoretically, this list should cover all the known species of animals. However, in practice this is infeasible. Moreover, the size of the list directly impacts the output of the ASR system. On the one hand, a missing keyword in the list will never be recognized; on the other hand, a longer list will increase the perplexity of the keyword model. The automatic construction of an adequate language model for this exercise was done by using an existing extensive resource and then reducing its size by weighting the terms with a probability according to its popularity (returned by a web search engine) and imposing a threshold value. Uncommon results were also filtered with Onto.PT\(^1\), an ontology for the Portuguese language with reduced coverage [7]. Nevertheless, the generated results revealed an uneven distribution of probabilities, and so the list was divided in 10 classes that compose a histogram of words with all the terms in a class having the same probability and decreasing this probability proportionally with the class order. Finally, in order to balance the weight of the competing background model to the keywords, several configurations were exploited to adjust the background penalty to the best compromise. As a final result, we obtained a list with 802 animal names, weighted with probabilities. This list was then used to generate the language model for the Animal Naming exercise upon the stimuli creation.

Contrarily to the other exercises, for the evocation exercises the system is not supposed to consider the words in the list of right answers as synonyms, but instead the platform should consider them as individual words that are scored 1 if recalled by the patient. Thus, the maximum score corresponds to the number of words that compose the list, except for the Animal Naming test for which there is no maximum.

5 Evaluation

In order to evaluate the screening tool which was the subject of the present thesis, we need to collect speech data both from patients and from healthy people. However, it is infeasible for a person to perform all the neuropsychological

\(^1\) http://ontopt.dei.uc.pt/ (visited in May 2014)
tests implemented at once, as this would result in a too long session (mainly because of the extensiveness of the ADAS-cog test). The solution was to select a representative subset of these tests and combine them in a single global test developed only for recording purposes.

The speech corpus includes recordings from 10 senior subjects, 5 female and 5 male. All subjects were native Portuguese, of which 5 are healthy and 5 were diagnosed with cognitive impairment. No particular constraint over background noise condition was imposed. Each session consisted approximately of a 20 to 30-minutes recording. The data was originally captured with the platform at 16 kHz, and later down-sampled to 8 kHz to match the acoustic models sampling frequency.

In order to evaluate the keyword spotting (KWS) exercises we used 2 different metrics: Word Verification Rate (WVR) and Root-Mean-Square Error (RMSE). First, the WVR metric was applied to verify the reliability of the platform as a verification tool. Then, the RMSE metric was used to measure the differences between the scores computed manually and automatically. The evocation exercises were evaluated with 2 different metrics: Word Error Rate (WER) and RMSE. Regarding the WER measure, it was used to evaluate the performance of the ASR system for this type of exercises, by calculating the error of the alignment between a reference (manual) and an hypothesis (automatic). Since our main goal is to verify the similarity between the automatic and the manual scores, only the RMSE results are presented here, in Table 1.

<table>
<thead>
<tr>
<th>Exercise type / Test</th>
<th>Patients</th>
<th>Healthy</th>
</tr>
</thead>
<tbody>
<tr>
<td>KWS (normal)</td>
<td>3.52</td>
<td>2.72</td>
</tr>
<tr>
<td>KWS (rule-based)</td>
<td>3.22</td>
<td>1.67</td>
</tr>
<tr>
<td>Evocation (except animals)</td>
<td>0.89</td>
<td>1.41</td>
</tr>
<tr>
<td>Animal Naming</td>
<td>3.38</td>
<td>2.14</td>
</tr>
<tr>
<td>MMSE</td>
<td>2.49</td>
<td>3.10</td>
</tr>
</tbody>
</table>

Table 1. RMSE by type of exercise and by neuropsychological test.

In general, the achieved results were better for healthy people than for patients. This is an expected result due to the impaired condition of the patients, which may manifest problems in terms of speech and a higher number of hesitations. This difficults the recognition process. Moreover, they may change their opinion when trying to answer, which is not so often for healthy people. Such situations may confuse the ASR system, specially for rule-based keyword spotting exercises. This is justified by the added complexity of the language models. Also, the noisy conditions were more pronounced in the place where the patient’s recordings were performed.

Overall, the obtained results suggest that the platform may be useful and reliable as a screening tool. However, the Animal Naming test still needs to be improved.
6 Conclusions and Future Work

This work provided an automatic web-based tool with SLT integration which may be used for screening purposes regarding cognitive impairments, with a main focus on AD. As far as we know, it is the only platform of this type implemented for the Portuguese population. The platform automatizes a set of neuropsychological tests that are commonly applied by therapists to assess the cognitive condition of a person. Particularly, a partial implementation of the MMSE and ADAS-cog tests was implemented, as well as the Animal Naming test, performing a total of 185 stimuli which comprise different challenges and solutions.

Since the screening tool resulted from an adaptation of another existing platform, i.e. VITHEA, previous work was needed before starting the concrete development process. Concretely, an additional effort was required to understand the existing system.

Regarding the evaluation process, it was necessary to test the platform both from patients and healthy people. The collection of the patients data, besides being emotionally demanding, it is a valuable resource which implied logistic difficulties.

We believe that this platform could help in the early diagnosis process of the disease, which is extremely important to slow down its progress. In fact, the evaluation revealed quite promising results, suggesting that this could be an added value for society. However, after performing tests with real patients, we noticed that an advanced impaired condition may difficult the use of the system. Furthermore, the condition of cognitive impairment combined with the ageing factor, often associated with deafness and currently with computer illiteracy, it is problematic. In fact, for patients with a more pronounced cognitive impairment or with auditory impairments, there may be difficulty in understanding the question being asked. Anyway, we may consider this platform as a preparation for the future once the future elderly generation will be computer literate.

Nevertheless, we expect that this tool will have adhesion for its usefulness and relevance. In fact, during test application, both the patients and healthy people demonstrated their appreciation for the platform, indicating that this is an interesting and appealing system. Even more, they demonstrated their interest in repeating the tests and using the platform regularly.

As future work the Animal Naming test should be improved by using different techniques to enhance the quality of the language model. Furthermore, new semantic categories may be also implemented. Ideally, the automatic generation of any semantic category would be of great interest.

Another desired improvement is to remove completely the mouse interaction with the platform during the test application. Currently, the action to stop the recording is still needed, as with the current methodology used is impossible to send the recording audio without the user interaction. The idea would be to automatically detect when to stop the recording through silence detection technique. As the exercise currently already advances on its own when the recording
is stopped, implementing this modification would enable to perform a complete
test without any interaction, in a fluid way with no interruptions.

Although we have slowed the speech rate of the virtual therapist to cope with
deafness problems often associated with the ageing factor, this was only a slight
improvement. Thus, as a future Master Thesis, an equivalent screening tool fully
adapted for people with deafness may be developed.

Lastly, a mobile version of the screening tool may also be developed as a
future Master Thesis, for example for tablets. This is motivated by the fact that
a touchscreen interface would simplify the interaction between the patient and
the system once this is more natural than using a mouse. Besides, this kind of
interface is also adequate if we intend to enable a more interactive platform by
implementing cognitive exercises which require manual actions, such as drawing.

References

1. C. Amorim. Automatic tool for screening of cognitive impairments. Master's thesis,
Instituto Superior Técnico, June 2014.
2. J. Barreto. Os sinais da doença e a sua evolução. In A. Castro-Caldas and A. Men-
donça, editors, A Doença de Alzheimer e Outras Demências em Portugal. LIDEL, 2005.
3. R. Coulston, E. Klabbers, J. Villiers, and J. Hosom. Application of speech technol-
ogy in a home based assessment kiosk for early detection of alzheimer’s disease. In
4. M. Guerreiro. Avaliação neuropsicológica das doenças degenerativas. In A. Castro-
Caldas and A. Mendonça, editors, A Doença de Alzheimer e Outras Demências em
Portugal. LIDEL, 2005.
5. P. Moreira and C. Oliveira. Fisiopatologia da doença de alzheimer e de outras
demências. In A. Castro-Caldas and A. Mendonça, editors, A Doença de Alzheimer
A Doença de Alzheimer e Outras Demências em Portugal. LIDEL, 2005.
7. H. Oliveira, L. Pérez, and P. Gomes. Integrating lexical-semantic knowledge to build
a public lexical ontology for portuguese. In Proc. 17th International Conference on
Applications of Natural Language Processing to Information Systems (NLDB), pages
In A. Castro-Caldas and A. Mendonça, editors, A Doença de Alzheimer e Outras
Demências em Portugal. LIDEL, 2005.