Multimedia therapeutic tool for Portuguese aphasics

Yves Rybarczyk[#1], Ricardo Martins[#2], José Fonseca[*3]

[#]*Department of Electrical Engineering, New University of Lisbon*
Quinta da Torre 2829-516 Monte de Caparica, Portugal
[1]y.rybarczyk@fct.unl.pt
[2]ricardomartins46@hotmail.com

[*]*Language Research Laboratory, Medical Faculty of Lisbon*
Av. Prof. Egas Moniz 1649-028 Lisboa, Portugal
[3]jfonseca@fm.ul.pt

***Abstract*— The system described in this article is designed for the treatment of the lusophone population of aphasic patients. Since this syndrome affects different functional modalities (written and spoken comprehension and expression) an implementation of a multimedia software has been necessary. Thus, the clinical material – developed by speech and language therapists – is displayed in the form of auditive, writing or virtual environments stimuli, following their relevance to the chosen task. Exercise results integrate multimodal contents and are essentially recorded into XML files, for a better flexibility of the data processing.**

***Keywords*— Stroke rehabilitation technology, serious games, 3D virtual environment, Human-Machine Interaction (HMI).**

I. Introduction

A. Characteristics of aphasia

Aphasia is the loss or impairment of language functions, such as oral expression, auditory comprehension, reading and writing. It is caused by brain damage. Aphasia disorders that impair motor programming (involved in expression and writing) are localized in anterior brain areas. On the other hand, posterior lesions generate auditive and reading comprehension impairments [1, for a review]. An aphasic syndrome can be more or less debilitating. In an initial stage some patients can only complete simple tasks of object identification, whereas other ones just need to exercise the comprehension of complex sentences. Although the tool described here is useful in both situations, it is particularly relevant for the most severe cases.

B. Rehabilitation techniques

1) Traditional approach: Multiple studies show speech therapy efficacy on aphasia recovery [2]. In this kind of treatment, the speech-language therapist stimulates the patient with questions, which the patient must answer. Indeed, an important factor in aphasia rehabilitation is the repetition of task completion [3]. Pencil and paper tests and bidimensional pictures (which illustrate objects and actions) are the most common materials used in recovery process. However, these rudimentary and relative representations of the physical world limit the evaluation of the patient's real speaking and comprehension skills. For instance, it is known that the visual discrimination of objects is more difficult when it is based on a two dimensional than on a normal three dimensions view, especially in persons with a low educational level [4].

2) IT approach: Since the 80's, the implementation of software as a form of aphasia treatment has been tried, using programs with a single as well as global language purpose [5] [6]. Several studies show how efficient IT is on aphasia treatment [7]. However, their diffusion has not been an easy process. Indeed, the IT developed until now has the disadvantages of being too expensive and using equipments too specialized to be used on a large scale and outside the clinical structure. Moreover, none of them allows a global treatment of the different aphasic syndromes, because they do not implement the auditive modality and/or display stimuli with a too basic grammatical construction [8].

The goal of this article is to describe a new generation of multimedia platforms for the aphasia treatment. In order to take into account the inter-individual differences of each clinical case, our tool enables the patient to train on the four modalities implicated in the aphasic syndrome, which are: reading, spoken understanding, spoken expression and writing. Furthermore, a particular caution was taken in the human-machine interface to ensure that it must be adapted to a large number of users (indeed, the syndrome is often correlated with other cerebral strokes that affect, particularly, the motor areas) and that it uses relatively ordinary materials to avoid a significant increase the platform cost. Finally, a huge spectrum of exercises of different complexity degrees was implemented in order to follow the progressive recovering of the patient. The results in each modality are almost exclusively recorded in a XML format, for a better management of the subsequent data processing.

II. System Design

A. Context of use

The objective of the software developed here is to improve the rehabilitation process of aphasic patients through a user-friendly virtual environment which is adapted to the everyday life needs of the users. Moreover, it ensures the recording and managing of a large number of information regarding the performance obtained in the various sessions of training, in order to carry out a future statistical analysis of the data. Although the patient could use the tool in an independent way, the assistance of a third party can be necessary, at least for the first sessions of practice. To notice that the tool is designed to be used at home (under the supervision of a relative) as well as in a clinical structure (under the orientation of a language therapist). This last condition is facilitated by the presence of an interface that enables the recording of different patients in a database. After his/her registration, the individual can initiate a task in one of the four symptomatic modalities of the aphasia: written comprehension, spoken comprehension, written expression or spoken expression. Each of the four themes is subdivided into sub-themes in relation with the modality in question. From the clinical material's point of view, the software loads multimedia information such as 3D immersive environments, video clips, voice synthesis and written stimuli.

B. Written comprehension

This modality is composed by five types of exercise. Three of them are displayed in the form of a tridimensional modeling of the home rooms: "noun matching", "sentence matching" and "sentence yes/no". The other two are only presented in the form of written items: "responsive naming" and "word intruder".

1) From a 3D environment: For these exercises the user has to choose the room where s/he wants to train (Fig. 1). Once the room is selected, s/he can use a control device (e.g., a joystick) to move inside the tridimensional environment in order to complete the orders that are written on the screen. In the case of a "noun matching", an object's name is displayed and the individual has to point/click on the object in question (Fig. 2). For a "sentence matching", the object to be found out is suggested by a sentence that defines this object. For instance, if the order "The place where we sit down for watching television" appears, the patient must point/click on the "sofa". At last, for a "sentence yes/no" task, a phrase is displayed on the screen and the user has to click on the "yes" or "no" following the affirmation, which is right or wrong in relation to the reality of the room. An example of possible sentence is: "The window is open…".

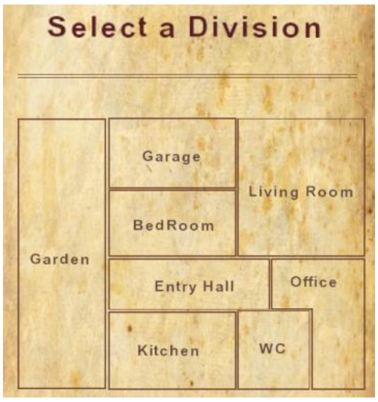

Fig. 1 Choice of the room

Also, the patient can use visual or/and audio helps when s/he has some difficulties to carry out the requested task. These supports are accessible through an icon and they consist of a picture of the object, when it is a visual cue, and the spelling of the object's name, when it is a phonetic cue. At the end of the stimuli set that composes an exercise, the performance of the individual is displayed on the computer screen, as well as the quantity of helps used.

Fig. 2 3D representation of the kitchen

2) From a written text: For a "responsive naming" task, an incomplete sentence and four possible words appear on the screen (Fig. 3). The patient has to choose one of the four words in order to complete the sentence in a logical way. Finally, for the research of the "word intruder", five words are displayed. Only four words have a semantic relationship (Fig. 4). The user has to click on the word that is not semantically related to the other ones. In these tasks as well as in the previous ones, the patient can ask for visual and/or audio cues to help him/her solve the exercise.

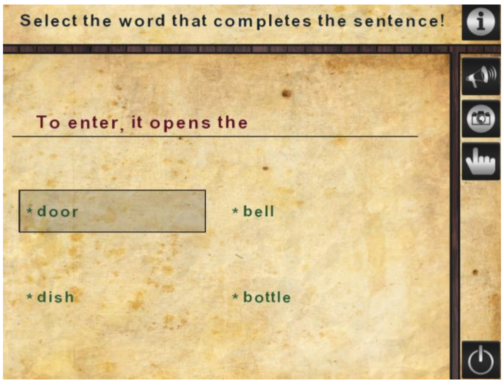

Fig. 3 Responsive naming (helps on the right)

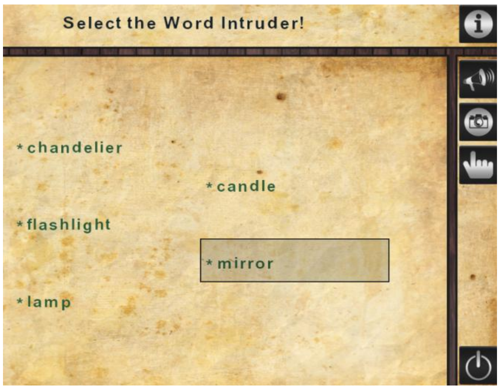

Fig. 4 Word intruder (helps on the right)

C. Spoken comprehension

Aphasia can also affect the auditive modality. In order to deal with this dimension, a second thematic is dedicated to the oral comprehension of the patient. This theme is divided into the three following sub-themes: “objects identification”, “sentence yes/no” and “simple orders”. The exercise set which is displayed in this condition is relatively similar with the one of the previous section. The main difference is that tasks are no more based on written stimuli but on phonetic ones instead (voice synthesis). Here, all exercises are carried out in a virtual environment of one of the home rooms the user selected. During the “object identification”, the patient listens to an object’s name and s/he has to point/click on the place where the object is. In a “sentence yes/no” task, the user is asked to answer “yes” or “no” to a sentence displayed in the form of a voice synthesis, depending if the affirmation is right or wrong. At last, the “simple orders” are audio sentences that indirectly refer to an object the patient must locate in the room. An example of simple order is: “When we open it, the water flows”. Helps are also available in order to provide, if necessary, visual and/or written cues to assist the patient in the task.

D. Written expression

So far, it is the most advanced theme of our tool. It is based on a clinical material implemented in multimedia information systems. In short, the stimuli are represented by a set of three kinds of content: “3D models”, “video clips” and “written information”.

1) From a tridimensional object: An object of the everyday life, which is modeled in 3D, is displayed on the screen. The patient has to write, through the computer keyboard, the name of the object in question (Fig. 5). A large set of supports is accessible to the user. For instance, s/he can listen to the object’s name or see a picture of it. From a word spelling point of view, some cues indicate the number of characters of the word and other cues show the next letter that composes the word. For individuals who need more assistance, an option allows to display three possibilities from which only one is the correct answer.

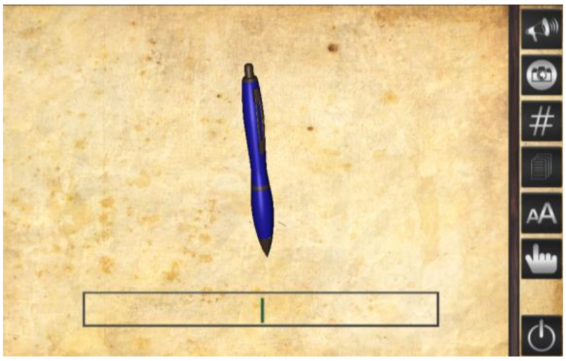

Fig. 5 To write a noun. On the right, the icons that represent the helps provided to the user. So, from the top to the bottom, the respective assistances are: to listen to the object’s name, to see the object’s picture, the number of characters that spells the name, three possible names, the next letter that composes the name, the expected answer and the option to close the session

2) From a video clip: A video clip of people carrying out an action is displayed on the screen. The patient’s task is to write the verb that matches this action. Thus, if the footage shows an individual who is dialing a telephone number, the user must write the verb “to phone”. Some scenes can be misunderstanding because of the subjectivity of the action. In these situations, the modeling of the data processing enables to consider different answers as true.

3) From a written information: Here, the user has to complete/restructure, in the correct form, a sentence or a whole paragraph. In the case of a sentence completion, the exercise can consist of introducing a noun, a conjugated verb or a preposition. In this kind of task, a sentence with a gap appears on screen and the patient has to fill it in. In regards to the sentence reconstruction, the task can involve i) a correction of phonologic errors (some words are badly spelled) or ii) a phrase rewriting through a dislocation of some words in order to construct a semantically correct sentence. Finally, more complex tasks require the patient to introduce words of different types (nouns, verbs, adjectives, …) in a text as large as a paragraph.

E. Spoken expression

This modality has a more complex development. Indeed, to work in a completely autonomous way, it needs a voice recognition program. However, despite fundamental progress in this area, voice recognition is still difficult to implement and faces reliability limitations [9]. This is the reason why the version of the platform presented in this article only allows a recording of the user’s spoken response, for further analysis. All the same, two kinds of exercise are currently implemented in the software: “objects naming” and “actions naming”. In the first case, the patient must pronounce, in the microphone, the name of the 3D object displayed on the screen. In the other case, s/he has to orally express the verb of the action s/he watches in a video clip.

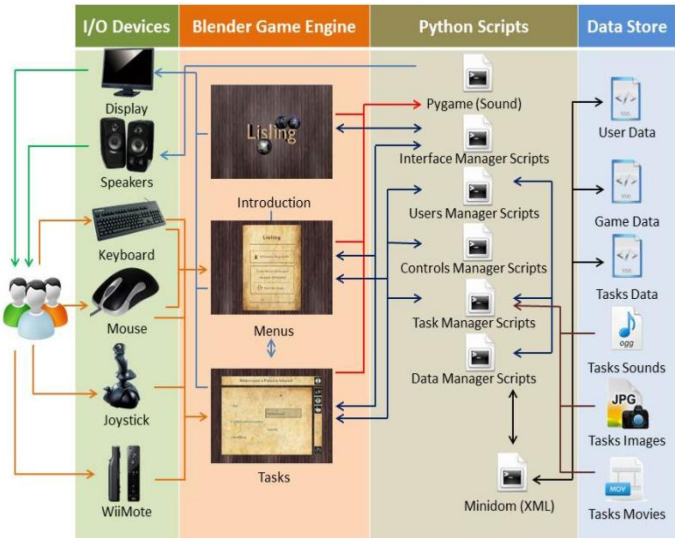

Fig. 6 Layers of the platform architecture

III. Software Architecture

The architecture of the multimedia platform is organized in these four layers: input/output devices, game engine, Python scripts and data management system (Fig. 6).

A. I/O devices

The fact that motor disabilities can usually be associated with an aphasic syndrome imposes an implementation of control interfaces, which are adapted to a large number of patients. These commands have to be as natural and intuitive as possible, because even patients without any experience in computers must be able to use the platform. In addition, the material cost was also a test of choice. Thus, besides the traditional input systems existing in a common computer (keyboard and mouse), a joystick and a Wii command were interfaced with the software. On one hand, the joystick is particularly useful to make the navigation inside the 3D environment easier. On the other hand, the Wii command may provide an upgrade in terms of degrees of freedom of movement, without loss of pointing precision. In return, this last device needs more configurations to connect with a personal computer (adaptation of the infrared sensor bar). In terms of outputs, the computer screen and speakers respectively enable to display visual and audio clinical materials which are used in the therapy.

B. Blender game engine

The object modeling, which is used to build the 3D environment of the home rooms, is carried out through Blender open source software. This platform takes the advantage to gather modeling, animation, video edition, scenes composition, interactive applications and game engine in a unique tool. Blender game engine allows, particularly, detecting collisions and reproducing physical properties such as gravity, friction forces, objects deformation… which are useful for a realistic simulation of the interactions with the tridimensional models. For our application, the game engine is also implicated in the creation/management of users, menus and specific files of each task.

C. Python scripts

Although Blender game engine is very rich, programming of additional Python scripts was necessary in order to add specific functionalities of the rehabilitation platform. These scripts are involved in the following operations:

- Menus management (linkage between Blender files and the scenes that compose them),
- Users management (creation/updating of the users and the recording of their respective data),
- Task management (loading of the data related to the exercises, checking of the responses and presentation of additional information such as actions and requested helps),
- Control management (listening of the inputs from the different control devices and translation of these values into specific actions in regards to the exercise in question).

D. Data management

Three kinds of data are created by the application: user data, session data and task data. User data are related to the identification, the personal information and preferences of each recorded patient.

```
<SESSION>
  <patient_id>2</patient_id>
  <session_id>5</session_id>
  <TASK>
    <task_id>3</task_id>
    <task_name>Written Comprehension</task_name>
    <exercise_name>Sentence Yes/No</exercise_name>
    <theme>Kitchen</theme>
```

Fig. 7 XML structure of the session data for the <SESSION> (identification of the user and session number) and <TASK> (task identification) tags

The session data collect all the actions and tasks carried out by a patient. The information is recorded in form of a XML file. Fig. 7 shows the main elements recorded under the <TASK> tag, which are the task name, the sub-task name and the room where the exercise is executed. The <OBJECTIVE> block is related to each objective requested to a user (in Fig. 8, the goal is to locate a “glace”). For each objective, the user can give various answers, which can be right or wrong. These attempts are recorded in the <ATTEMPT> structure (in Fig. 8, because the patient’s response is “glace”, the recorded result is “correct”). At last, the <HELP> structure saves all helps used by the patient (in the example below, the user asked to display a picture of the searched object).

```
    <OBJECTIVE>
      <objective_id>glace</objective_id>
      <ATTEMPT>
        <att_response>glace</att_response>
        <att_result>correct</att_result>
        <HELP>
          <help_name>picture</help_name>
        </HELP>
      </ATTEMPT>
    </OBJECTIVE>
  </TASK>
</SESSION>
```

Fig. 8 XML structure of the session data for the <OBJECTIVE> (e.g., researched object), <ATTEMPT> (patient’s response) and <HELP> (asked helps) tags

There is one XML file for each task that is implemented in the software. These files contain the data – which are previously inserted into the program – of all the objectives and their respective answers. Thanks to this architecture it is easy to insert new objectives, since we just have to add the structure corresponding to this objective to the XML file of task management. The next paragraphs describe typical examples of these files for three different tasks.

Sentence yes/no. The file of this kind of exercise is common for all tasks of written and spoken comprehension. Each objective is inserted into a structure that is related to a home room (in the example presented in Fig. 9, this is the living room). A specific tag (<question_id>) identifies the objective, which enables afterwards to cross-check the data between the responses the user gave and the answers that are expected by the system.

```xml
<LIVING_ROOM>
  <YESNO>
    <question_id>5</question_id>
    <question_text>Does the window is open?</question_text>
    <question_sound>sn5.ogg</question_sound>
    <answer>No</answer>
  </YESNO>
</LIVING_ROOM>
```

Fig. 9 XML structure of a "Sentence yes/no" task file. The <question_text> and <question_sound> tags respectively enable to display the question in written or spoken form. The <answer> tag indicates what the correct answer is

To write a verb. The patient has to write a verb in the infinitive matching the action displayed in a video clip. The <ANSWER> structure lists a set of different possible answers (the XML structure in Fig. 10 refers to a clip where an individual opens the door to another one). This task may also be carried out through a multiple choice exercise.

```xml
<MOVIE>
  <id>3</id>
  <film>vi3.flv</film>
  <ANSWER>
    <response>open</response>
    <response>welcome</response>
    <response>greet</response>
  </ANSWER>
  <multiple_choices>close,open,hit</multiple_choices>
</MOVIE>
```

Fig. 10 XML structure of a "To write a verb" task file

To write an object's name. The user has to write the name of the object that is presented in the form of a 3D model. Once again, various responses are acceptable. Fig. 11 shows the XML structure for the pen object (<obj_name> and <obj_sound> for the respective identification of the loaded model and sound). Different kinds of supports are available, such as multiple proposals (<help_multiple>), the sound of the first syllable (<help_1_sound>), the first syllable of the word (<help_1_sil>), the number of characters (help_n_letters>) or the object's picture (<help_picture>). The <ANSWERS> block lists all the responses that are considered as correct.

```xml
<OBJECT>
  <obj_id>2</obj_id>
  <obj_name>pen</obj_name>
  <obj_sound>pen.ogg</obj_sound>
  <help_multiple>pen,tooth-brush,knife</help_multiple>
  <help_1_sound>pen_1.ogg</help_1_sound>
  <help_1_sil>pe</help_1_sil>
  <help_n_letters>***</help_n_letters>
  <help_picture>pen.jpeg</help_picture>
  <ANSWERS>
    <ans>pen</ans>
    <ans>Biro</ans>
    <ans>ball-point</ans>
    <ans>fountain pen</ans>
  </ANSWERS>
</OBJECT>
```

Fig. 11 XML structure of a "To write an object's name" task file

IV. Conclusions

The software developed ensures a complete linguistic rehabilitation of aphasic patients, in the sense that it enables the user to practice both grammatical and functional aspects of the language and to train all the modalities implicated in the linguistic communication (spoken and written). A game engine is used to display the clinical exercises through a 2D and 3D environment. This serious game approach is chosen to enhance the exercise frequency – thanks to a playful platform – which is the main factor implicated in aphasic recovering [10] [11]. Various human-machine interfaces are available in order to be customizable according to each case of aphasic symptoms – with the possible brain damages that can also affect the patient's motor control – and to be adapted to the different kind of exercises. For a better flexibility of the data processing, the information is structured in the form of XML files. We take advantage of this format to make the integration of new contents easier, in order to create a dynamic tool. The objective is to allow the platform customization through the possibility that the user will have to load some personal items. The next step will be to test the software in an aphasic population to evaluate the relevance of session training on the patient's daily activities, and to compare the performance with a traditional speech therapy.

References

[1] H. Damasio, *Neuroanatomical correlates of the aphasias.* Acquired Aphasia. San Diego, CA: Academic Press, 1991.

[2] A. Basso, *Aphasia and its Therapy*. New York, NY: Oxford University Press, 2003.

[3] S.K. Bhogal, R. Teasell, and M. Speechley, "Intensity of aphasia therapy, impact on recovery", *Stroke*, vol. 34, pp. 987-993, 2003.

[4] A. Reis, K.M. Petersson, A. Castro-Caldas, and M. Ingvar, "Formal schooling influences two-but not three-dimensional naming skills", *Brain and Cognition*, vol. 47, pp. 397-411, 2001.

[5] Y. Rybarczyk, and J. Fonseca, "Tangible interface for a rehabilitation of comprehension in aphasic patients", in *Proc. AAATE'11,* 2011, Maastricht, Netherlands.

[6] J. Fonseca, G. Leal, L. Farrajota, J. Carneiro, and A. Castro-Caldas, "Lisling – Multimedia program for aphasics training", in *Proc. AAATE'97*, 1997, Chalkidike, Greece.

[7] I. Robertson, "Does computerized cognitive rehabilitation work?", *Aphasiology*, vol. 4, pp. 381-405, 1990.

[8] C. Sik Lányi, Z. Geiszt, and V. Magyar, "Using IT to inform and rehabilitate aphasic patients", *Informing Science Journal*, vol. 9, pp. 163-179, 2006.

[9] D. Pogue, "Talk to the machine", *Scientific American*, vol. 303, p. 18, 2010.

[10] L.R. Cherney, A.S. Halper, A.L. Holland, J.B. Lee, E, Babbitt, and R. Cole, "Improving conversational script production in aphasia with virtual therapist computer treatment software", *Brain and Language*, vol. 103, pp. 246-247, 2007.

[11] W.M.E. Van de Sandt-Koenderman, "Aphasia rehabilitation and the role of computer technology: can we keep up with modern times", *International Journal of Speech-Language Pathology*, vol. 13, pp. 21-27, 2011.