An on-line system for remote treatment of aphasia

Anna Pompili, Alberto Abad, Isabel Trancoso
L^{2}F - Spoken Language Systems Lab
INESC-ID/IST, Lisbon, Portugal
{anna,alberto,imt}@l2f.inesc-id.pt

José Fonseca, Isabel P. Martins, Gabriela Leal, Luisa Farrajota
LEL - Language Research Laboratory
Lisbon Faculty of Medicine, Portugal
jfonseca@fm.ul.pt

Abstract

Aphasia treatment for the recovery of lost communication functionalities is possible through frequent and intense speech therapy sessions. In this sense, speech and language technology may provide important support in improving the recovery process. The aim of the project Vithea (Virtual Therapist for Aphasia Treatment) is to develop an on-line system designed to behave as a virtual therapist, guiding the patient in performing training exercises in a simple and intuitive fashion. In this paper, the fundamental components of the Vithea system are presented, with particular emphasis on the speech recognition module. Furthermore, we report encouraging automatic word naming recognition results using data collected from speech therapy sessions.

1 Introduction

Aphasia is a communication disorder that can affect various aspects of language, including hearing comprehension, speech production, and reading and writing fluency. It is caused by damage to one or more of the language areas of the brain. Many times the cause of the brain injury is a cerebral vascular accident (CVA), but other causes can be brain tumors, brain infections and severe head injury due to an accident. Unfortunately, in the last decades the number of individuals that suffer CVAs has dramatically increased, with an estimated 600.000 new cases each year in the EU. Typically, a third of these cases present language deficiencies (Pedersen et al., 1995). This kind of language disorder involves countless professional, family and economic problems, both from the point of view of the individual and the society. In this context, two remarkable considerations have led to the development of the Portuguese national project Vithea (Virtual Therapist for Aphasia treatment).

First are the enormous benefits that speech and language technology (SLT) may bring to the daily lives of people with physical impairment. Information access and environment control are two areas where SLT has been beneficially applied, but SLT also has great potential for diagnosis, assessment and treatment of several speech disorders (Hawley et al., 2005). For instance, a method for speech intelligibility assessment using both automatic speech recognition and prosodic analysis is proposed in (Maier et al., 2009). This method is applied to the study of patients that have suffered a laryngotomy and to children with cleft lip and palate. (Castillo-Guerra and Lovey, 2003) presents a method for dysarthria assessment using features extracted from pathological speech signals. In (Yin et al., 2009), the authors describe an approach to pronunciation verification for a speech therapy application.

The second reason for undertaking the Vithea project is that several aphasia studies have demonstrated the positive effect of speech therapy activities for the improvement of social communication abilities. These have focused on specific linguistic impairments at the phonemic, semantic or syntactic levels (Basso, 1992). In fact, it is believed more and more that the intensity of speech therapy positively influences language recovery in aphasic patients (Bhogal et al., 2003).

These compelling reasons have motivated the de-

Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies, pages 1–10,
Edinburgh, Scotland, UK, July 30, 2011.

velopment of an on-line system for the treatment of aphasic patients incorporating recent advances in speech and language technology in Portuguese. The system will act as a "virtual therapist", simulating an ordinary speech therapy session, where by means of the use of automatic speech recognition (ASR) technology, the virtual therapist will be able to recognize what was said by the patient and to validate if it was correct or not. As a result of this novel and specialized stimulation method for the treatment of aphasia, patients will have access to word naming exercises from their homes at any time, which will certainly cause an increase in the number of training hours, and consequently it has the potential to bring significant improvements to the rehabilitation process.

In section 2 we provide a brief description of different aphasia syndromes, provide an overview of the most commonly adopted therapies for aphasia, and describe the therapeutic focus of our system. Section 3 is devoted to an in depth description of the functionalities that make up the system, while section 4 aims at detailing its architecture. Finally, section 5 describes the automatic speech recognition module and discusses the results achieved within the automatic naming recognition task.

2 About the aphasia disorder

2.1 Classification of aphasia

It is possible to distinguish two different types of aphasia on the basis of the fluency of the speech produced: fluent and non-fluent aphasia. The speech of someone with fluent aphasia has normal articulation and rhythm, but is deficient in meaning. Typically, there are word-finding problems that most affect nouns and picturable action words. Non-fluent aphasic speech is slow and labored, with short utterance length. The flow of speech is more or less impaired at the levels of speech initiation, the finding and sequencing of articulatory movements, and the production of grammatical sequences. Speech is choppy, interrupted, and awkwardly articulated.

Difficulty of recalling words or names is the most common language disorder presented by aphasic individuals (whether fluent or non-fluent). In fact, it can be the only residual defect after rehabilitation of aphasia (Wilshire and Coslett, 2000).

2.2 Common therapeutic approaches

There are several therapeutic approaches for the treatment of the various syndromes of aphasia. Often these methods are focused on treating a specific disorder caused from aphasia. The most commonly used techniques are output focused, such as the stimulation method and the Melodical Intonation Therapy (MIT) (Albert et al., 1994). Other methods are linguistic-oriented learning approaches, such as the lexical-semantic therapy or the mapping technique for the treatment of agrammatism. Still, several non-verbal methods for the treatment of some severe cases of non-fluent aphasia, such as the visual analog communication, iconic communication, visual action and drawing therapies, are currently used (Sarno, 1981; Albert, 1998).

Although there is an extensive list of treatments specifically designed to recover from particular disorders caused by aphasia, one class of rehabilitation therapy especially important aims to improve the recovery from word retrieval problems, given the widespread difficulty of recalling words or names. Naming ability problems are typically treated with semantic exercises like *Naming Objects* or *Naming common actions* (Adlam et al., 2006). The approach typically followed is to subject the patient to a set of exercises comprising a set of stimuli in a variety of tasks. The stimuli are chosen based on their semantic content. The patient is asked to name the subject that has been shown.

2.3 Therapeutic focus of the Vithea system

The focus of the Vithea system is on the recovery of word naming ability for aphasic patients. So far, experiments have only been made with fluent aphasia patients, but even for this type of aphasia, major differences may be found. Particularly, patients with Transcortical sensorial aphasia, Conduction aphasia and Anomic aphasia (Goodglass, 1993) have been included in our studies.

Although the system has been specifically designed for aphasia treatment, it may be easily adapted to the treatment or diagnosis of other disorders in speech production. In fact, two of the patients that have participated in our experimental study were diagnosed with acquired apraxia of speech (AOS), which typically results from a stroke,

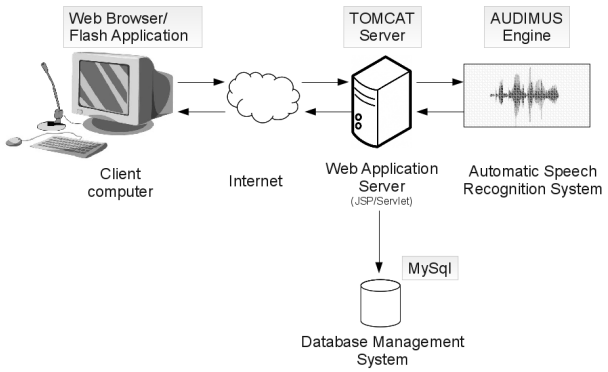

Figure 1: Comprehensive overview of the Vithea system.

tumor, or other known neurological illness or injury, and is characterized by inconsistent articulatory errors, groping oral movements to locate the correct articulatory position, and increasing errors with increasing word and phrase length.

3 The Vithea System

The overall flow of the system can be described as follows: when a therapy session starts, the virtual therapist will show to the patient, one at a time, a series of visual or auditory stimuli. The patient is required to respond verbally to these stimuli by naming the contents of the object or action that is represented. The utterance produced is recorded, encoded and sent via network to the server side. Here, a web application server receives the audio file via a servlet that serves as an interface to the ASR system, which takes as input the audio file encoding the patient's answer and generates a textual representation of it. This result is then compared with a set of predetermined textual answers (for that given question, of course) in order to verify the correctness of the patient's input. Finally, feedback is sent back to the patient. Figure 1 shows a comprehensive view of this process.

The system comprises two specific modules, dedicated respectively to the patients for carrying out the therapy sessions and to the clinicians for the administration of the functionalities related to them. The two modules adhere to different requirements that have been defined for the particular class of user for which they have been developed. Nonetheless they share the set of training exercises, that are built by the clinicians and performed by the patients.

3.1 Speech therapy exercises

Following the common therapeutic approach for treatment of word finding difficulties, a training exercise is composed of several semantic stimuli items. The stimuli may be of several different types: text, audio, image and video. Like in ordinary speech therapy sessions, the patient is asked to respond to the stimuli verbally, describing the imaging he/she sees or completing a popular saying (which was presented verbally or in text).

Exercise categories

The set of therapeutic exercises integrated in Vithea has been designed by the Language Research Laboratory of the Department of Clinical Neuroscience of the Lisbon Faculty of Medicine (LEL). LEL has provided a rich battery of exercises that can be classified into two macro-categories according to the main modality of the stimulus, namely:

A) Image or video: *Naming object picture, Naming of verbs with action pictures, and Naming verbs given pictures of objects.*

B) Text or speech: *Responsive Naming, Complete Sayings, Part-whole Associations, What name is given to. . . , Generic Designation, Naming by function, Phonological Evocation, and Semantics Evocation.*

Exercises can be also classified according to *Themes*, in order to immerse the individual in a pragmatic, familiar environment: a) *Home b) Animals c) Tools d) Food e) Furniture f) Professions g) Appliances h) Transportation i) Alive/Not Alive j) Manipulable/Not Manipulable k) Clothing l) Random.*

Evaluation exercises

In addition to the set of training exercises, which are meant to be used on a daily basis by the aphasic patient, the Vithea system also supports a different class of exercises: Evaluation Exercises. Unlike training exercises, evaluation exercises are used by human therapists to periodically assess the patient's progress and his/her current degree of aphasia via an objective metric denoted as Aphasia Quotient (AQ). Evaluation exercises are chosen from a

subset of the previously mentioned classes of therapeutic exercises, namely: *Naming object picture*, *Naming of verbs with action pictures*, and *Naming verbs given pictures of objects*.

3.2 Patient Module

The patient module is meant to be used by aphasic individuals to perform the therapeutic exercises.

Visual design considerations

Most of the users for whom this module is intended have had a CVA. Because of this, they may have some forms of physical disabilities such as reduced arm mobility, and therefore they may experience problems using a mouse. Acknowledging this eventuality, particular attention has been given to the design of the graphical user interface (GUI) for this module, making it simple to use both at the presentation level and in terms of functionality provided. Driven by the principle of accessibility, we designed the layout in an easy to use and understand fashion, such that the interaction should be predictable and unmistakable.

Moreover, even though aphasia is increasing in the youngest age groups, it still remains a predominant disorder among elderly people. This age group is prone to suffer from visual impairments. Thus, we carefully considered the graphic elements chosen, using big icons for representing our interface elements. Figure 2 illustrates some screenshots of the Patient Module on the top.

Exercise protocol

Once logged into the system, the virtual therapist guides the patient in carrying out the training sessions, providing a list of possible exercises to be performed. When the patient choses to start a training exercise, the system will present target stimuli one at a time in a random way. After the evaluation of the patient's answer by the system, the patient can listen again to his/her previous answer, record again an utterance (up to a number of times chosen before starting the exercise) or pass to the next exercise.

Patient tracking

Besides permitting training sessions, the patient module has the responsibility of storing statistical and historical data related to user sessions. User utterances and information about each user access to the system are stored in a relational database. Particularly, start and end time of the whole training session, of a training exercise, and of each stimulus are collected. On the one hand, we log every access in order to evaluate the impact and effectiveness of the program by seeing the frequency with which it is used. On the other hand, we record the total time needed to accomplish a single stimulus or to end a whole exercise in order to estimate user performance improvements.

3.3 Clinician Module

The clinician module is specifically designed to allow clinicians to manage patient data, to regulate the creation of new stimuli and the alteration of the existing ones, and to monitor user performance in terms of frequency of access to the system and user progress. The module is composed by three submodules: **User, Exercise, Statistic**.

User sub-module

This module allows the management of a knowledge base of patients. Besides basic information related to the user personal profile, the database also stores for each individual his/her type of aphasia, his/her aphasia severity (7-level subjective scale) and AQ information.

Exercise sub-module

This module allows the clinician to create, update, preview and delete stimuli from an exercise. An exercise is composed of a varying number of stimuli. In addition to the canonical valid answer, the system accepts for each stimulus an extended word list comprising three extra valid answers. This list allows the system to consider the most frequent synonyms and diminutives.

Since the stimuli are associated with a wide assortment of multimedia files, besides the management of the set of stimuli, the sub-module also provides a rich Web based interface to manage the database of multimedia resources used within the stimuli. Figure 2c shows a screenshot listing some multimedia files. From this list, it is possible to select a desired file in order to edit or delete it.

In this context, a preview feature has also been provided. The system is capable of handling a wide range of multimedia encoding: audio (accepted file

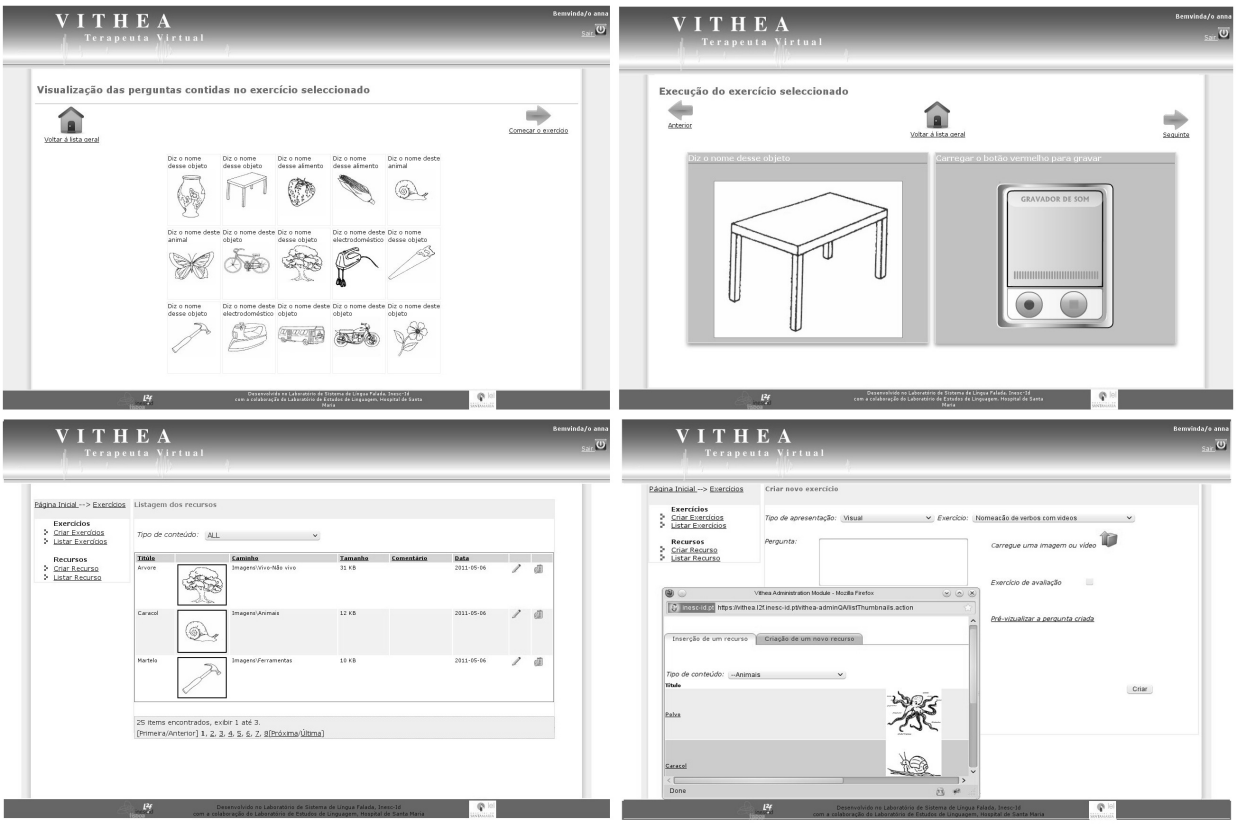

Figure 2: Vithea system screenshots: a) Interface with preview of the stimuli constituting an exercise in the patients module (top-left), b) interface for performing a specific stimulus in the patients module (top-right), c) interface for the management of multimedia resources in the clinician module (bottom-left) and d) interface for the creation of new stimulus in the clinician module (bottom-right).

types: *wav, mp3*), video (accepted file types: *wmv, avi, mov, mp4, mpe, mpeg, mpg, swf*), and images (accepted file types: *jpe, jpeg, jpg, png, gif, bmp, tif, tiff*).

Given the diversity of the various file types accepted by the system, a conversion to a unique file type was needed, in order to show them all with only one external tool. Audio files are therefore converted to *mp3* file format, while video files are converted to *flv* file format.

Finally, a custom functionality has been designed to create new stimuli in an intuitive fashion similar in style to a WYSIWYG editor. Figure 2d illustrates the stimuli editor, showing how to insert a multimedia resource.

Statistics sub-module

This module allows the clinician both to monitor statistical information related to user-system interactions and to access the utterances produced by the patient during the therapeutic sessions. The statistical information comprises data related to the user's progress and to the frequency with which users access the system. On the one hand, we provide all the attempts recorded by the patients in order to allow a re-evaluation by clinicians. This data can be used to identify possible weaknesses or errors from the recognition engine. On the other hand, we thought that monitoring the utilization of the application from the users could be an important piece of feedback about the system's feasibility. This is motivated by common concerns about the fact that some users abandon their therapeutic sessions when they are not able to see quick results in terms of improvements.

4 Architectural Overview

Considering the aforementioned requirements and features that will make up the system, Learning Management Systems (LMSs) software applications were initially considered. LMSs automate the ad-

ministration of training events, manage the log-in of registered users, manage course catalog, record data from learners and provide reports to the management (Aydin and Tirkes, 2010). Thus, an in-depth evaluation of the currently widespread solutions was carried out (Pompili, 2011). Concretely, eight different LMSs (Atutor, Chamilo, Claroline, eFront, Ilias, Moodle, Olat, Sakai) were studied in detail. Unfortunately, the outcome of this study revealed important drawbacks.

The main problem noticed is that LMSs are typically feature-rich tools that try to be of general purpose use, sometimes resulting in the loss of their usefulness to the average user. Often the initial user reaction to the interface of these tools is confusion: the most disorienting challenge is figuring out where to get the information needed. As previously mentioned, patients who have had a CVA may experience physical deficiencies, thus the Vithea system needs an easy to use and understandable interface. We dedicated some effort trying to personalize LMS solutions, but most of them do not allow easy simplification of the presentation layout.

Moreover, while there were several differences between the functionalities that the evaluated LMSs provided in terms of training exercises, they all presented various limitations in their implementation. Eventually, we had to acknowledge that it would have been extremely complex to customize the evaluated frameworks to meet the Vithea project requirements without introducing major structural changes to the code.

Besides, the average user for whom the Vithea system is intended is not necessarily accustomed with computers and even less with these tools, which in most cases are developed for environments such as universities or huge organizations. This means that our users may lack the technical skills necessary to work with an LMS, and the extra effort of understanding the system would result in a loss of motivation.

Therefore, considering the conclusions from this study, we have opted to build a modular, portable application which will totally adhere to our requirements. With these purposes in mind, the system has been designed as a multi-tier web application, being accessible everywhere from a web browser. The implementation of the whole system has been achieved by integrating different technologies of a heterogeneous nature. In fact, the presentation tier exploits Adobe®Flash®technology in order to support rich multimedia interaction. The middle tier comprises the integration of our own speech recognition system, AUDIMUS, and some of the most advanced open source frameworks for the development of web applications, Apache Tiles, Apache Struts 2, Hibernate and Spring. In the data tier, the persistence of the application data is delegated to the relational database MySQL. This is where the system maintains information related to patient clinical data, utterances produced during therapeutic sessions, training exercises, stimuli and statistical data related both to the frequency with which the system is used, and to the patient progress.

4.1 Speech-related components of the system

Audio Recorder

In order to record the patient's utterances, the Vithea system takes advantage of opportunities offered by Adobe®Flash®technology. This allows easy integration in most browsers without any required extra plugin, while avoiding the need for security certificates to attest to the reliability of an external component running in the client machine within the browser. This choice was mainly motivated from the particular kind of users who will use the system, allowing them to enjoy the advantages of the virtual therapist without the frustration of additional configuration. A customized component has been developed following the aforementioned principles of usability in terms of designing the user interface. Keeping simplicity and understandability as our main guidelines, we used a reduced set of large symbols and we tried to keep the number of interactions required to a bare minimum. Therefore, recording and sending an utterance to the server requires only that the patient starts the recording when ready, and then stops it when finished. Another action is required to play back the recorded audio.

Automatic Speech Recognition Engine

AUDIMUS is the Automatic Speech Recognition engine integrated into the Vithea system. The AUDIMUS framework has been developed during the last years of research at the Spoken Language Processing Lab of INESC-ID (L^2F), it has been success-

fully used for the development of several ASR applications such as the recognition of Broadcast News (BN) (Meinedo et al., 2010). It represents an essential building block, being the component in charge of receiving the patient answers and validating the correctness of the utterances with respect to the therapeutic exercises. In the following section, this specific module of the Vithea architecture is assessed and described in more detail.

5 The Vithea speech recognition module

5.1 The AUDIMUS hybrid speech recognizer

AUDIMUS is a hybrid recognizer that follows the connectionist approach (Boulard and Morgan, 1993; Boulard and Morgan, 1994). It combines the temporal modeling capacity of Hidden Markov Models (HMMs) with the pattern discriminative classification of multilayer perceptrons (MLP). A Markov process is used to model the basic temporal nature of the speech signal, while an artificial neural network is used to estimate posterior phone probabilities given the acoustic data at each frame. Each MLP is trained on distinct feature sets resulting from different feature extraction processes, namely Perceptual Linear Predictive (PLP), log-RelAtive SpecTrAl PLP (RASTA-PLP) and Modulation SpectroGram (MSG).

The AUDIMUS decoder is based on the Weighted Finite State Transducer (WFST) approach to large vocabulary speech recognition (Mohri et al., 2002).

The current version of AUDIMUS for the European Portuguese language uses an acoustic model trained with 57 hours of downsampled Broadcast News data and 58 hours of mixed fixed-telephone and mobile-telephone data (Abad and Neto, 2008).

5.2 Word Naming Recognition task

We refer to *word recognition* as the task that performs the evaluation of the utterances spoken by the patients, in a similar way to the role of the therapist in a rehabilitation session. This task represents the main challenge addressed by the virtual therapist system. Its difficulty is related to the utterances produced by aphasic individuals that are frequently interleaved with disfluencies like hesitation, repetitions, and doubts. In order to choose the best approach to accomplish this critical task, preliminary evaluations were performed with two sub-sets of the Portuguese Speech Dat II corpus. These consist of word spotting phrases using embedded keywords: the development set is composed of 3334 utterances, while the evaluation set comprises 481 utterances. The number of keywords is 27. Two different approaches were compared: the first based on large vocabulary continuous speech recognition (LVCSR), the second based on the acoustic matching of speech with keyword models in contrast to a background model. Experimental results showed promising performance indicators by the latter approach, both in terms of Equal Error Rate (EER), False Alarm (FA) and False Rejection (FR). Thus, on the basis of these outcomes, background modeling based keyword spotting (KWS) was considered more appropriate for this task.

Background modeling based KWS

In this work, an equally-likely 1-gram model formed by the possible target keywords and a competing background model is used for word detection. While keyword models are described by their sequence of phonetic units provided by an automatic grapheme-to-phoneme module, the problem of background modeling must be specifically addressed. The most common method consists of building a new phoneme classification network that in addition to the conventional phoneme set, also models the posterior probability of a background unit representing "general speech". This is usually done by using all the training speech as positive examples for background modeling and requires retraining the acoustic networks. Alternatively, the posterior probability of the background unit can be estimated based on the posterior probabilities of the other phones (Pinto et al., 2007). We followed the second approach, estimating the posterior probability of a garbage unit as the mean probability of the top-6 most likely outputs of the phonetic network at each time frame. In this way there is no need for acoustic network re-training. Then, a likelihood-dependent decision threshold (determined with telephonic data for development) is used to prune the best recognition hypotheses to a reduced set of sentences where the target keyword is searched for.

5.3 Experiments with real data

Corpus of aphasic speech

A reduced speech corpus composed of data collected during therapy sessions of eight different patients has been used to assess the performance of the speech recognition module. As explained above, two of them (patients 2 and 7) were diagnosed with AOS. Each of the sessions consists of naming exercises with 103 objects per patient. Each object is shown with an interval of 15 seconds from the previous. The objects and the presentation order are the same for all patients. Word-level transcription and segmentation were manually produced for the patient excerpts in each session, totaling 996 segments. The complete evaluation corpus has a duration of approximately 1 hour and 20 minutes.

Evaluation criteria

A word naming exercise is considered to be completed correctly whenever the targeted word is said by the patient (independently of its position, amount of silence before the valid answer, etc...). It is worth noticing that this is not necessarily the criterion followed in therapy tests by speech therapists. In fact, doubts, repetitions, corrections, approximation strategies and other similar factors are usually considered unacceptable in word naming tests, since their presence is an indicator of speech pathologies. However, for the sake of comparability between a human speech therapist evaluation and an automatic evaluation, we keep this simplified evaluation criterion. In addition to the canonical valid answer to every exercise, an extended word list containing the most frequent synonyms and diminutives has been defined, for a total KWS vocabulary of 252 words. Only answers included in this list have been accepted as correct in both manual and automatic evaluation.

Results

Word naming scores are calculated for each speaker as the number of positive word detections divided by the total number of exercises (leftmost plot of Figure 3). The correlation between the human evaluation assessed during ordinary therapeutic sessions and the automatic evaluation assessed with the word recognition task has resulted in a Person's coefficient of 0.9043. This result is considered quite promising in terms of global evaluation. As concerning individual evaluations (rightmost plot of Figure 3), it can be seen that the system shows remarkable performance variability in terms of false alarms and misses depending on the specific patient. In this sense, the adaptation to the specific user profile may be interesting in terms of adjusting the system's operation point to the type and level of aphasia. As a preliminary attempt to tackle the customization issue, the word detector has been individually calibrated for each speaker following a 5-fold cross-validation strategy with the corresponding patient exercises. The calibration is optimized to the minimum false alarm operation point for patients with high false-alarm rates (2, 3, 4, 5 and 8) and to the minimum miss rate for patients with a high number of misses (1, 6 and 7). Figure 4 shows results for this customized detector. In this case, the correlation between human and automatic evaluation is 0.9652 and a more balanced performance (in terms of false alarm and false rejection ratios) is observed for most speakers.

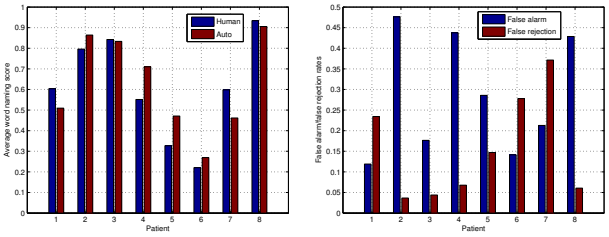

Figure 3: *On the left side, average word naming scores of the human and automatic evaluations. On the right side, false alarm and false rejection rates.*

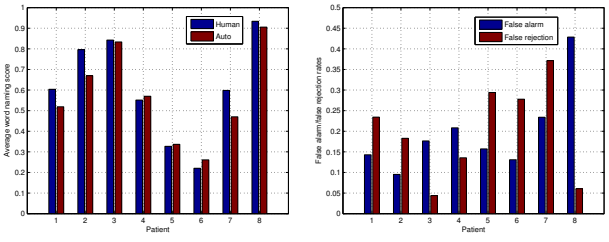

Figure 4: *On the left side, average word naming scores of the human and automatic evaluations with the customized detector. On the right side, false alarm and false rejection rates of the customized detector.*

Analysis of word detection errors

The most common cause for false alarms is the presence of many "invented" nonexistent words without semantic meaning, which are very often phonetically very close to the target words. These paraphasic errors were present in all types of fluent aphasia and AOS that we have observed, but not for all patients. In many of these errors, the stressed syllable is often pronounced right, or at least its rhyme. As the typical stress pattern in Portuguese is in the penultimate syllable, most often the last syllable is also pronounced correctly (e.g. borco / porco). In patients that try to say the word by approximation, that is, by successive attempts to get closer to the target word, but using only existent words, the differences between the percentages of miss and false alarms are not so remarkable.

One characteristic of aphasic patients that sometimes causes keywords to be missed (both when correctly or incorrectly pronounced) is pauses in between syllables. This may justify the inclusion of alternative pronunciations, in case such pronunciations are considered acceptable by therapists. Additionally, more sophisticated speech tools may also be integrated, such as tools for computing the goodness of pronunciation (Witt, 1999). This would allow a different type of assessment of the pronunciation errors, which may provide useful feedback for the therapist and the patients.

6 Conclusions and future work

6.1 Conclusions

This paper described how automatic speech recognition technology has contributed to build up a system that will act as a virtual therapist, being capable of facilitating the recovery of people who have a particular language disorder: aphasia. Early experiments conducted to evaluate ASR performance with speech from aphasic patients yielded quite promising results.

The virtual therapist has been designed following relevant accessibility principles tailored to the particular category of users targeted by the system. Special attention has been devoted to the user interface design: web page layout and graphical elements have been chosen keeping in mind the possibility that a user may experience reduced arm mobility and the technology that has been integrated was selected with the idea of minimizing possible difficulties in using the system. A pedagogical approach has been followed in planning the functionalities of the virtual therapist. This has been mainly driven by the fundamental idea of avoiding an extra feature rich tool which could have resulted in frustration for some patients, who seek help for recovery and do not need to learn how to use complex software.

Overall, since the system is a web application, it allows therapy sessions anywhere at anytime. Thus, we expect that this will bring significant improvements to the quality of life of the patients allowing more frequent, intense rehabilitation sessions and thus a faster recovery.

6.2 Future work

The Vithea system has recently achieved the first phase of a project which still entails several improvements. Even though, *Naming objects* and *Naming common actions* are the most commonly used exercises during the rehabilitation therapies, the system has been designed to allow a more comprehensive set of therapeutic exercises which will be implemented during the next refinement phase. Also, at this stage, we plan to make available the current version of the system to real patients in order to receive effective feedback on the system.

In the subsequent improvement phase, we will integrate the possibility of providing help, both semantic and phonological to the patient whenever the virtual therapist is asked for. Hints could be given both in the form of a written solution or as a speech synthesized production based on Text To Speech (TTS). Furthermore, we are considering the possibility of incorporating an intelligent animated agent that together with the exploitation of synthesized speech, will behave like a sensitive and effective clinician, providing positive encouragements to the user.

Acknowledgements

This work was funded by the FCT project RIPD/ADA/109646/2009, and partially supported by FCT (INESC-ID multiannual funding) through the PIDDAC Program funds. The authors would like to thank to Prof. Dr. M. T. Pazienza, A. Costa and the reviewers for their precious comments.

References

A. Abad and J. P. Neto. 2008. International Conference on Computational Processing of Portuguese Language, Portugal. *Automatic classification and transcription of telephone speech in radio broadcast data*.

A. L. R. Adlam, K. Patterson, T. T. Rogers, P. J. Nestor, C. H. Salmond, J. Acosta-Cabronero and J. R. Hodges. 2006. Brain. *Semantic dementia and Primary Progressive Aphasia: two side of the same coin?*, 129:3066–3080.

M. L. Albert, R. Sparks and N. A. Helm. 1994. Neurology. *Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Assesment: melodic intonation therapy*, 44:566–568.

M. L. Albert. 1998. Arch Neurol-Chicago *Treatment of aphasia*, 55:1417–1419.

C. C. Aydin and G. Tirkes. 2010. Education Engineering. *Open source learning management systems in e-learning and Moodle*, 54:593–600.

A. Basso. 1992. Aphasiology. *Prognostic factors in aphasia*, 6(4):337–348.

S. K. Bhogal, R. Teasell and M. Speechley. 2003. Stroke. *Intensity of aphasia therapy, impact on recovery*, 34:987–993.

H. Bourlard and N. Morgan. 1993. IEEE Transactions on Neural Networks. *Continuous speech recognition by connectionist statistical methods*, 4(6):893–909.

H. Bourlard and N. Morgan. 1994. Springer. *Connectionist speech recognition: a hybrid approach*.

D. Caseiro, I. Trancoso, C. Viana and M. Barros. 2003. International Congress of Phonetic Sciences, Barcelona, Spain. *A Comparative Description of GtoP Modules for Portuguese and Mirandese Using Finite State Transducers*.

E. Castillo-Guerra and D. F. Lovey. 2003. 25th Annual Conference IEEE Engineering in Medicine and Biology Society. *A Modern Approach to Dysarthria Classification*.

H. Goodglass. 1993. *Understanding aphasia: technical report*. Academy Press, University of Califórnia. San Diego.

M. S. Hawley, P. D. Green, P. Enderby, S. P. Cunningham and R. K. Moore. 2005. Interspeech. *Speech technology for e-inclusion of people with physical disabilities and disordered speech*, 445–448.

A. Maier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster and E. Nöth. 2009. Speech Communication. *PEAKS - A System for the Automatic Evaluation of Voice and Speech Disorders*, 51(5):425–437.

H. Meinedo and J. P. Neto. 2000. International Conference on Spoken Language Processing, Beijing, China. *Combination Of Acoustic Models In Continuous Speech Recognition Hybrid Systems*, 2:931–934.

H. Meinedo, A. Abad, T. Pellegrini, I. Trancoso and J. P. Neto. 2010. Fala 2010, Vigo, Spain. *The L2F Broadcast News Speech Recognition System*.

M. Mohri, F. Pereira and M. Riley. 2002. Computer Speech and Language. *Weighted Finite-State Transducers in Speech Recognition*, 16:69–88.

P. M. Pedersen, H. S. Jørgensen, H. Nakayama, H. O. Raaschou and T. S. Olsen. 1995. Ann Neurol *Aphasia in acute stroke: incidence, determinants, and recovery*, 38(4):659–666.

J. Pinto, A. Lovitt and H. Hermansky. 2007. Interspeech. *Exploiting Phoneme Similarities in Hybrid HMM-ANN Keyword Spotting*, 1817–1820.

A. Pompili. 2011. Thesis, Department of Computer Science, University of Rome. *Virtual therapist for aphasia treatment*.

M. T. Sarno. 1981. *Recovery and rehabilitation in aphasia*, 485–530. Acquired Aphasia, Academic Press, New York.

C. E. Wilshire and H. B. Coslett. 2000. *Disorders of word retrieval in aphasia theories and potential applications*, 82–107. Aphasia and Language: Theory to practice, The Guilford Press, New York.

S. M. Witt. 1999. *Use of speech recognition in Computer assisted Language Learning*. PhD thesis, Department of Engineering, University of Cambridge.

S. -C. Yin, R. Rose, O. Saz and E. Lleida. 2009. IEEE International Conference on Acoustics, Speech and Signal Processing. *A study of pronunciation verification in a speech therapy application*, 4609–4612.