VITHEA: On-line therapy for aphasic patients exploiting automatic speech recognition

Anna Pompili and Alberto Abad

L²F - Spoken Language Systems Lab, INESC-ID Lisboa, Portugal
{anna.pompili,alberto.abad}@l2f.inesc-id.pt
http://www.l2f.inesc-id.pt/

Abstract. Aphasia is an acquired communication disorder that affects speech and language functionalities at varying degrees. The recovery of lost communication functionalities is possible through frequent and intense speech therapy sessions. The aim of the VITHEA -Virtual Therapist for Aphasia Treatment- project is to exploit speech and language technology (SLT) to facilitate the recovery process of Portuguese aphasic patients, more concretely to recover lost word naming abilities. The proposed system has been designed to behave as a virtual therapist that simulates ordinary speech therapy sessions and, by means of automatic speech recognition (ASR) technology, validates patients performance.

Keywords: speech disorder, aphasia, speech recognition, keyword spotting, word naming, virtual therapy, on-line recovery

1 Introduction

Aphasia is a communication disorder caused by the damage of one or more language areas of the brain affecting various speech and language functionalities, including hearing comprehension, speech production, and reading and writing fluency. Most of the times aphasia can be fully recovered, however, one common residual disorder after rehabilitation among aphasia patients is the difficulty to recall names or words. Typically, word retrieval problems can be treated through semantic exercises like *Naming Objects or Naming common actions* [1]. The approach typically followed is to subject the patient to a set of stimuli in a variety of tasks. Frequency and intensity of these therapy sessions has been proven to be a key factor in the recovery of lost communication functionalities [2].

VITHEA is the first prototype of an on-line platform that incorporates SLT for treatment of Portuguese speakers with aphasia [3]. The system aims at acting as a "virtual therapist", like in ordinary speech therapy sessions it asks the patient to recall the contents represented in a photo or picture and by means of the use of ASR technology processes what is said by the patient and decides either if it is correct or wrong. Thus, the platform makes available word naming exercises to aphasia patients from their homes at any time, which will certainly cause an increase in the number of training hours, and hopefully a significant improvement in the rehabilitation process.

2   VITHEA: The proposed system

VITHEA is a web-based platform that permits speech therapists to easily create speech therapy exercises that can be later accessed by aphasia patients using a web-browser. During the training sessions, the role of the speech therapist is taken by a "virtual therapist" that presents the exercises and that is able to validate the patients answers.

The overall flow of the system can be described as follows: when a therapy session starts, the virtual therapist shows to the patient, one at a time, a series of visual or auditory stimuli. The patient is then 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 and it is processed by 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 the given question) 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.
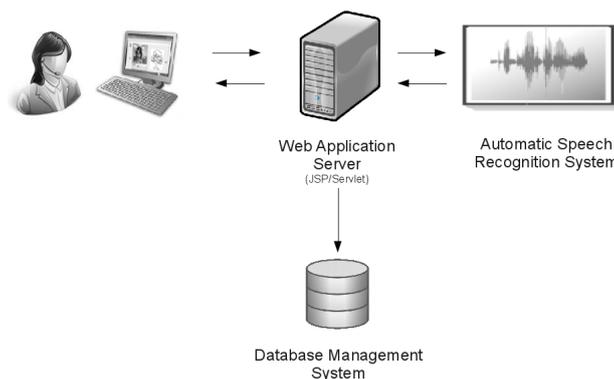

Fig. 1. Comprehensive overview of the VITHEA system.

Automatic Speech Recognition The ASR module is the backbone of the system: it is responsible for receiving the patient's speech answer and validating the correctness of the utterance for a given therapeutic exercise. Consequently, it strongly determines the usability of the whole therapeutic platform. The targeted task for automatic word naming recognition consists of deciding whether a claimed word W is uttered in a given speech segment S or not. This task is referred to as word verification, although we actually perform term verification, since a keyword may in fact consist of more than one word (e.g. *chapéu de chuva*). Several approaches exist based on speech recognition technology to tackle the word verification problem. Given that word W is known, forced alignment with an automatic speech recognition system could be one of the most straightforward possibilities. However, we expect that speech from aphasic patients will

contain a considerable amount of hesitations, doubts, repetitions, descriptions and other speech disturbing factors that can make this approach inconvenient. Alternatively, keyword spotting methods can better deal with unexpected speech effects. In fact, word verification can be considered a particular case of keyword spotting (with a single search term) and similar approaches can be used. In the current version of the system, acoustic based keyword spotting is applied for word verification. In order to do so, our in-house ASR engine named AUDIMUS, that has been previously used for the development of several ASR applications, was modified to incorporate a competing background speech model that is estimated without the need for acoustic model re-training similarly to [4].

The patient and clinician modules The system comprises two specific web applications: the patient and the clinician modules. They are dedicated respectively to the patients for carrying out the therapy sessions and to the clinicians for the administration of the functionalities related to them. Both 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. The patient exercise interface has been designed to cope with the functionalities needed for automatic word recalling therapy exercises, which includes besides the integration of an animated virtual character (the virtual therapist), text-to-speech synthesized voice, image and video displaying, speech recording and playback functionalities, automatic word naming recognition and exercise validation and feed-back prompting.

3   Demo details

During the demo session, the patient and clinician modules will be shown. Since both modules are remote web applications, one or more computers will be provided, although any computer with a built-in or external microphone, Internet connection and a web-browser can be used to access the system. Portuguese speaking attendees will be able to directly test the system: for the patient module they will be able to experience word naming therapy exercises and for the clinician module they will be able to easily create and edit new exercises.

References

1. Adlam, A. L. R., Patterson, K., Rogers, T. T., Nestor, P. J., Salmond, C. H., Acosta-Cabronero J., Hodges, J. R.: Semantic dementia and Primary Progressive Aphasia: two side of the same coin? Brain. 129:3066–3080 (2006)
2. Basso, A.: Prognostic factors in aphasia. Aphasiology. 6(4):337–348 (1992)
3. Pompili, A., Abad, A., Trancoso, I., Fonseca, J., Martins, I., Leal, G., Farrajota, L.: An on-line system for remote treatment of aphasia. In: Proc. Second Workshop on Speech and Language Processing for Assistive Technologies (2011)
4. Pinto, J., Lovitt, A., Hermansky, H.: Exploiting phoneme similarities in hybrid hmm-ann keyword spotting. In: Proc. Interspeech. 2007, pp. 1610–1613 (2007)