Residential Automation with Vocal Interaction in Brazilian Portuguese Language

Artur Oliveira[1], Carlos Pimentel[1], Eduardo Silva[1], Gabriel Araújo[1], Giovanny Lucero[1], Hendrik Macedo[1], Leonardo Matos[1], Miguel Soto, Ricardo Seixas, Thiago Mendonça[1], Thiago Reis[1]

[1] Universidade Federal de Sergipe, Departamento de Computação, Cidade Universitária Prof. Aloísio Campos, Rosa Elze, São Cristóvão – SE, Brazil
{artur_aeee88@hotmail.com, fragapimentel@gmail.com, edseabrasilva@yahoo.com.br, gabrielfeear@hotmail.com, giovanny@ufs.br, hendrik@ufs.br, lnmatos@ufs.br, miguel.pari@gmail.com, ricardo.seixas@iade.com.br, thiagodm@dcomp.ufs.br, thiago.ufs@gmail.com}

Abstract. This paper presents a software for recognizing Brazilian Portuguese language in a context of residential automation. This software is an interface layer between the user and an application that controls the equipments in the house.

Keywords: Vocal Interaction, Residential Automation, Brazilian Portuguese, Speech Recognition.

1 Introduction

On the last decade, the concept of *"intelligent houses"* has been created, which are residences that feature embedded systems in their electronical devices, allowing residents to interact with them, automating and facilitating home tasks.

There are several ways of interacting with residential automation systems. One of them is vocal interaction. Therefore, the main goal for this work is to develop a speech recognition system for Brazilian Portuguese language in order that it can be used as a module for vocal interaction in residential automation.

2 System Architecture

In the front-end of the system, the speaker interacts with a Java application for the Android OS, through a smarthphone. The application captures the audio in *wav* format and sends to the server on a desktop. Another Java application in server side receives acoustic signals embedded in wave files and interacts with *Julius* engine. The engine performs recognition based on an acoustic and a language model, resulting the recognized phrase. In the last stage, the recognized sentence is analyzed by the ontology, which checks whether it agrees or not semantically with the scenario.

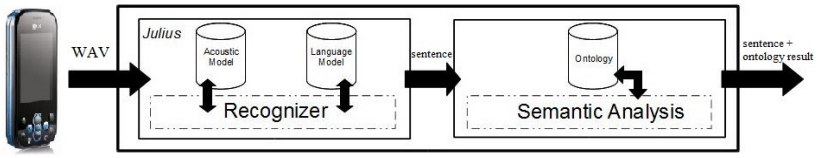

Fig. 1. System architecture.

The system architecture is illustrated on Figure 1. The acoustic model for the present work has been trained with the Spoltech *corpus*, using the *HTK* tool to create the Hidden Markov Models. The language model consists of a tri-gram model. To create the model, a *corpus* was built with ≃10000 sentences, with ≃ 200 words total.

In a first stage of the semantic evaluation of the sentence, the keywords are extracted and then sent to the ontology, responsible for the domain knowledge. The ontology then checks the correctness of the sentence. For example, if the recognizer outputs "turn on the microwave" or "turn microwave on", the ontology will understand that these are semantically correct, but if "on microwave turn" or "irrigate microwave" are given, they will be discarded. In the first case because the keywords order is incorrect. In the second case because it doesn't make sense for the domain.

3 Demo

In this demo, the software will be demonstrated running on a personal computer. While a person speaks on a microphone, the system recognizes it in real-time and outputs the recognized sentence, along with the ontology output. There will be no demonstration with a smartphone due to this feature is not yet fully implemented.

4 Results and Discussion

For testing the software, a base with 32 spoken sentences of a male speaker was used. The base is composed of 16kHz (mono, 16-bit) WAV files. These sentences consist of phrases used in the domain, some simple and other more complex. To consider that a sentence was correctly recognized, the output is compared with the actually spoken phrase. Slight variations are accepted, taking only keywords in consideration.

Using this criterion, from the 32 sentences, the recognizer got 27 sentences marked as correct (all verified to be consistent), what gives a 84,3% global recognition rate.

Due to the complexity of the Portuguese language, both in variety of phonemes and phrase construction, it is a challenge to achieve good recognition rates in such a task. The global recognition rates from the tests were considered good but not entirely satisfactory. Changes can be made to the software to refine the recognition.