IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 10, Issue 12, December 2020, ||Series -II|| PP 01-05

Prototype of a Voice-Controlled Home Automation System

Mauro Leonardo Pérez Bravo [1], Jaime Hugo Puebla Lomas[2]

[1] *Escuela Superior de Cómputo del Instituto Politécnico Nacional, CDMX, México*

[2] *Departamento de Ingeniería en Sistemas Computacionales, ESCOM-IPN, CDMX, México*

Received 15 December 2020; Accepted 29 December 2020

ABSTRACT: This work proposes to develop a home automation system (intelligent home control) with a voice-based user interface that does not depend on an external Internet service to perform speech processing. The application of the system takes place in scenarios where it is sought to automate tasks in the home of users who require a more natural interface, compared to the use of computer applications, tablet or smartphone, for example, elderly adults, blind or with a motor disability that prevents the use of such applications through a conventional device. The system consists of a prototype of wireless communication modules and an embedded system with software that allows processing a set of voice commands to interpret the actions and execute the corresponding tasks

KEYWORDS: Markov's hidden model, home automation, voice recognition, embedded systems, wireless nodes.

I. INTRODUCTION

Domotics is the science that studies the design and construction of automation and control systems, for different environments with multiple objectives. There are standards that define the construction and implementation standards for home automation systems, as well as various protocols that allow the system hardware to communicate. However, a standardized and universally accepted communication protocol is still being developed to allow the interoperability of existing products [1].

Speech-to-text is the name given to the technology that allows the processing of human speech and converting what the user says into text. There are currently many technological products that offer this service and in general the operating scheme consists of a mobile device (tablet or smartphone) capturing what the user says through a microphone and then sending the information over the Internet to be processed by each service provider and finally sending the response back to the device as text [2].

A voice-controlled home automation system is a control or automation system that uses a speech-to-text interface to allow the user to operate the system. Through the pronunciation of a set of spoken instructions, called voice commands, the user of the home automation system is able to instruct the system to execute the tasks for which it is built or programmed.

I.1 Problem Statement

Home automation is a growing trend that despite presenting new advances every day, still does not meet the requirements of some sectors of the population; speaking specifically of older adults, having a home automation system becomes a necessity for daily life and therefore requires an interface as natural as possible to interact with the system.

Likewise, as it becomes a system of vital importance for the life of certain sectors of the population, it is essential to reduce as much as possible the dependencies with other systems in order to avoid that the availability and functionality of the domotic system is reduced by failures or intermittencies in the services on which it depends.

Therefore, a system is sought that does not depend on an Internet Service Provider (ISP) as well as a speech-to-text service provider so that it is capable of performing voice processing locally and without the precondition of having a functional communications infrastructure (structured cabling or wireless network from equipment provided by an ISP).

I.2 Solution proposal

It is proposed to develop a prototype of a voice-controlled home automation system that is able to recognize a defined set of voice commands and operate without the dependency on an ISP. This system must be capable of performing all the processing involved in voice recognition and communication with the different parties locally and without requiring additional support infrastructure to the electrical network of a home.

Additionally, it must observe both the feature of easy installation and non-invasiveness for the user, as well as low power consumption.

I.3 Recognition models

Once the basic parameters for speech recognition are chosen and obtained, valid methods and algorithms are implemented that translate the numerical values from the signal processing stages into symbolic values that can be easily identified.

In the area of speech recognition, the sounds represented by spectral parameters are not easily distinguished by people, making the final stage of recognition more arduous, less intuitive and less given to the use of empirical methods than in other cases.

Due to the difficulty in using empirical recognition methods based on spectral parameters, the need to make use of automatic parametric recognition techniques arises, that is, automated techniques that discover the relationships between the numerical parameters obtained and the desired symbolic ones.

Among the best known and most widely used methods of parametric automatic recognition are Markov chains and neural networks, although there are others such as genetic algorithms. Markov chains are widely used in the industrial and commercial environment.

In general terms, the objective of a speech recognition system is to achieve automatic identification of some part of some type of spoken message. A relatively simple problem to solve today is the recognition of words that are assumed to have various types of restrictions, including a single speaker, reduced vocabulary, and absence of noise among the most common.

A common first step in the recognition of related speech is to try to separate each word from the text and, sometimes, representative sequences smaller than the words. This stage is called segmentation.

Segmentation is a very complex stage that cannot be fully resolved without resorting to the recognition phase. Voice recognisers try to isolate portions of speech as words, syllables, demialophones, diphonemes or basic sounds that can be studied in an isolated manner. These portions correspond to the segments and the complete process is segmentation.

In order to develop methods to obtain discriminant characteristics of the signals to be recognized, it is necessary to have a deep knowledge of the initial structures. In the case of spoken language, a good example of determining characteristics are formants.

The automatic parametric recognition models are characterized by the automatic way in which the spectral characteristics of the sounds are extracted. Once the representative sounds of speech are selected and organized in groups, they are presented to the parametric algorithms and, through supervised learning, the knowledge of the spectral characteristics of these sounds is internally represented in the internal structures that each model contains.

Once the learning phase is over, one can move on to the recognition phase, in which the patterns obtained from the speaker are confronted with the information of the internal states, and in the case of Markov chains they are easier to use than neural networks and do not present convergence problems that often appear in the latter, which can be implemented over a wide range of the language.

II.1 METHODOLOGICAL FRAME

For the implementation of this prototype, an adaptation of the V model for the development of embedded systems was taken into account. It consists of 7 stages, which start with an analysis and design, following an implementation and finally a debugging and final integration. The stages of this model are shown in Figure 1.

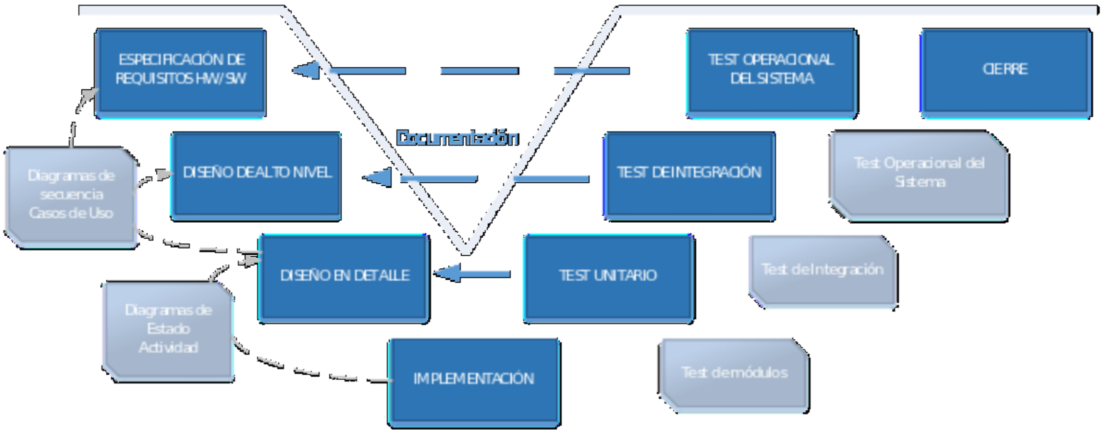

Figure 1. V Model

Starting from the requirements specification, it is intended to define and document the different requirements of the system to be implemented, following a global design which aims to obtain a general vision of the system.

The detailed design consists of specifying each block of the previous phase, here we intend to describe the design of the actuator node, the microphone node, the embedded system and speech recognition software, followed by the implementation of each of these.

The unit test checks each HW and SW module individually, where each of the modules will be debugged until the desired result is obtained. The integration phase joins the different modules of the system following the operational test, where the last tests are performed on a real scenario.

II.2 ANALYSIS AND DESIGN OF THE SYSTEM PROTOTYPE

The system consists of prototype hardware modules with wireless communication capability and an embedded system with software that allows processing a set of voice commands to interpret the actions and execute the tasks indicated. Specifically, the main components of the system are the following:

1. Microphone node: it is the hardware component that allows receiving the user's spoken instructions, digitizing them and sending them wirelessly to the embedded processing system.

2. Embedded system for speech processing and interpretation: it is the hardware component in charge of processing the digitized voice signal to convert it into text and thus evaluate the action that the user wants to execute, to send it wirelessly to the corresponding actuator node.

3. Actuator node. It is the hardware component that allows to control the turning on and off of different electronic devices in the home, as well as the lighting, by manipulating their switches wirelessly.

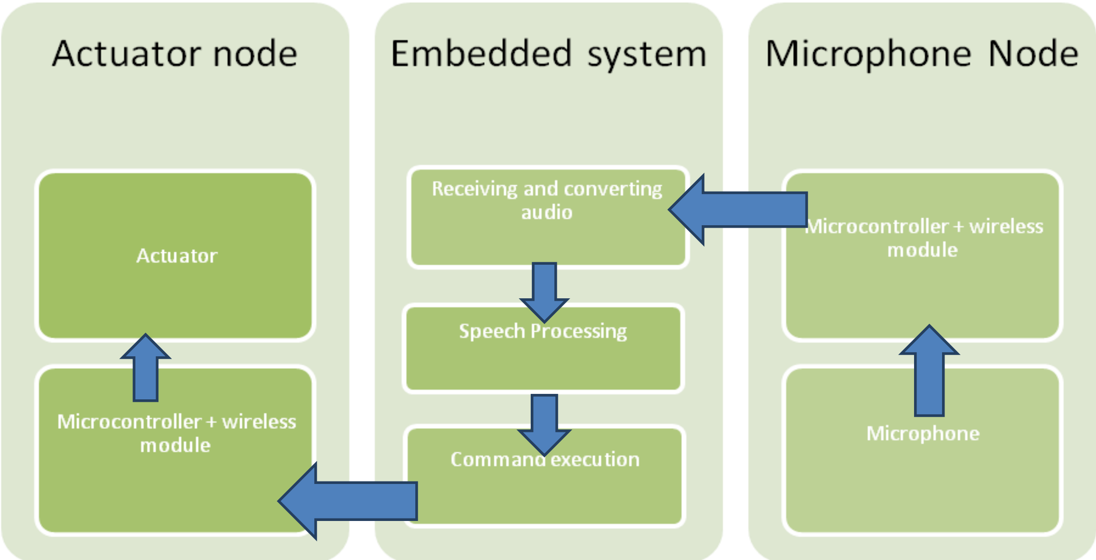

Figure 2. Prototype Block Diagram

II.2.1 SCHEMATIC DESIGN OF THE ELECTRONIC CIRCUIT

For the implementation of the actuator node uses a Zener diode 12V to 1W of power to feed the relay. Also includes a voltage regulator 3.3V because the module NRF24L01 + works with this voltage[3].

The figure 3, shows the interconnections between the wireless module and the microcontroller, in this implementation, the PIC12F509. The rectifier bridge has a capacity of 1A and for the PCB design it is used in a round package (7.6mm).

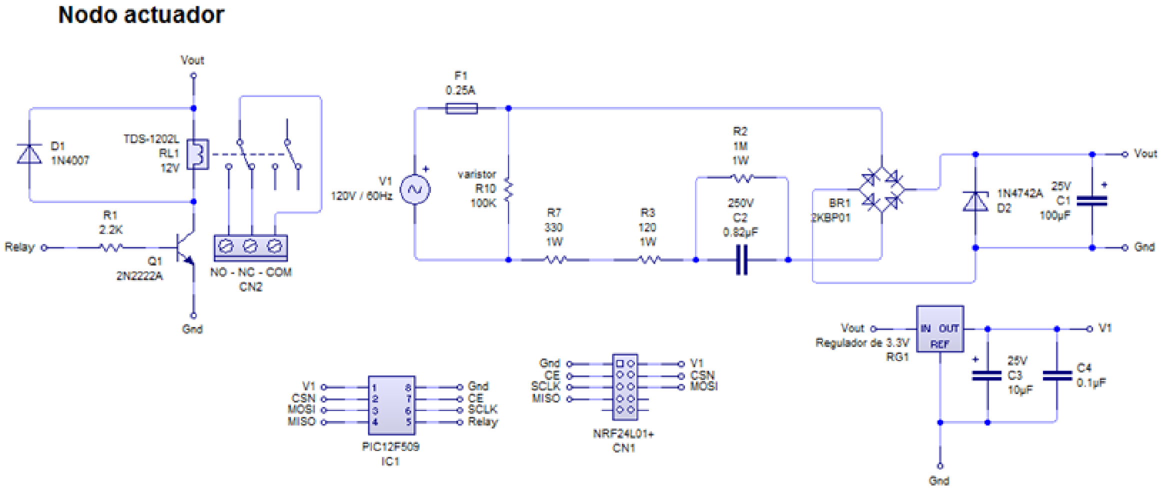

Figure 3. Electronic diagram of the actuator node

The 12V relay must work with a maximum coil current of 16.7mA in order to operate properly with the 21mA source and for the PCB a DPDT is used. The protection varistor must be 130V. The socket of the NRF24L01+ module is actually 8 positions. The figure 4 shows the top view of the manufactured node.

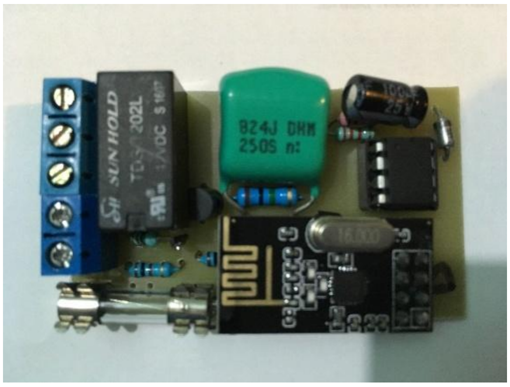

Figure 4. Prototype of the manufactured actuator node

II.2.2 EMBEDDED SYSTEM FOR PROCESSING VOICE COMMANDS

The embedded system considered in this prototype is implemented with a Raspberry Pi 3 model B. As for the software components implemented in this prototype, there are modules developed in C and Python language to perform the communication tasks with the communication devices that send voice signal samples and receive the instructions to operate the actuators and to perform the recognition of the system's voice commands.

II.2.2.1 RESULTS OF THE FASTDTW ALGORITHM

This section describes the results obtained with the FastDTW algorithm implemented in the embedded system using Python language[4]. The process consists of acquiring voice samples, pronouncing certain commands and separating the samples that will be used as reference patterns and those that will be assumed to be random samples to be identified.

For each sample a data file is obtained with the help of the software for communication with hardware devices developed in C, which is used to plot and display the signal and to build the WAV audio files. For the purposes of system prototyping, the FastDTW algorithm works with the sample data files. View Figure 5.

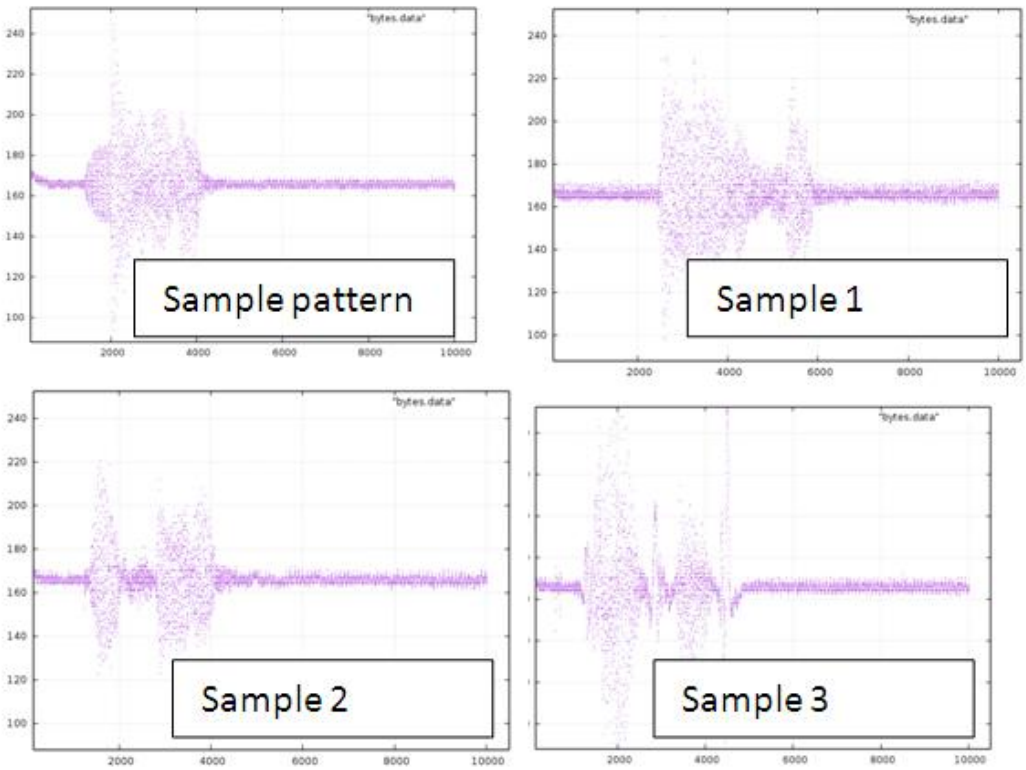

Figure 5. Sample of voice

III. CONCLUSIONS

The results of the development of this prototype are consistent with the general and particular objectives set at the beginning of the project. The objective of recognizing a limited set of voice instructions and processing them in an isolated embedded system (without internet connection) was achieved.

During the realization of the project, some areas of improvement were identified in relation to several components of the system. On the one hand, the actuator nodes present instability after a prolonged use (more than 8 hrs). For this reason, it is considered important to evaluate the redesign of the system to include better technology (solid state relays, for example).

Regarding the microphone node, the main challenge was to meet the speed requirements of the components so as not to affect the sampling frequency. Regarding this element, we also see the possibility of improving the dimensions of the PCB designed to maintain the desired characteristics.

The embedded speech recognition system presented a limited development compared to the original perspective of carrying out the full implementation of speech-to-text, however, the main requirement was met by implementing the FastDWT algorithm that made possible the measurement of similarity between speech signals.

REFERENCES

[1]. S. Heath, Embedded Systems Design, Massachusetts: Elsevier, 2002.
[2]. E. M. Pérez, Microcontroladores PIC: sistema integrado para el auto aprendizaje, Barcelona: Marcombo, 2007.
[3]. S.nordic, NRF24L01p.
[4]. E. Al. Fast DTW

Mauro Leonardo Pérez Bravo, et. al. "Prototype of a Voice-Controlled Home Automation System." *IOSR Journal of Engineering (IOSRJEN)*, 10(12), 2020, pp. 01-05.