International Journal of Scientific Engineering and Technology Research
www.ijsetr.com
ISSN 2319-8885
Vol.04,Issue.36,
September-2015,
Pages:7701-7708

Automatic Speech Recognition for Resource Constrained Embedded Systems

RAJULAPATI POORNA RAGHAVA RAO[1], KESANA MOHANA LAKSHMI[2]
[1]PG Scholar, Dept of ECE, CMR Technical Campus, Hyderabad, TS, India.
[2]Associate Professor, Dept of ECE, CMR Technical Campus, Hyderabad, TS, India.

Abstract: The design and implementation of prototype hardware/software architecture for automatic single word speech recognition on resource-constrained embedded devices Designed as a voice-activated extension of an existing wireless nurse call system, our prototype device continually listens for a pre-recorded keyword, and uses speech recognition techniques to trigger an alert upon detecting a match. Preliminary experiments show that our prototype achieves a high average detection rate of 96%, while only dissipating 28.5 mW for continuous audio sampling and duty-cycled speech recognition. The speech recognition is based on the accent of the English language used in this project which allows performing the operations based on the words spell. The words which are spell are stored in the operating system of Raspbian OS which is implemented on Raspberry Pi hardware kit of ARM 11 processor. The design is based on the architecture which determines the output according to the speech which is previously stored during operation of the system. The implementation is basically performed with the LED and fan which is turned ON & OFF based on voice recognition. This design can be used in real time implementation of the wide varied applications. Based on the voice recognition, the application tends to perform the changes of the output which can be used in the real time applications.

Keywords: Duchene Muscular Dystrophy (DMD), Low-Power Wireless Bus (LWB), Voice Activity Detection (VAD).

I. INTRODUCTION

Low-power wireless networks for healthcare are largely being used for monitoring patient activity and physiological parameters. Instead, we recently developed a wireless nurse call system, providing bidirectional interactions between patient and caregiver, based on push-buttons and off-the-shelf low-power wireless devices. We successfully deployed the system in a summer camp for boys with Duchene muscular dystrophy (DMD). Patient devices were installed at the beds so the boys could alert a caregiver for help during the night. Using the Low-Power Wireless Bus (LWB), our system provides highly reliable and timely bidirectional interactions between wireless patient devices and a centralized graphical user interface through which caregivers acknowledge alerts. Since boys with an advanced stage of DMD experience difficulty pressing a button, we have developed a voice activated extension to our wireless nurse call system. To make it convenient for a boy to request help, the new patient device must automatically detect a single word with high accuracy. Furthermore, to simplify deployment and ensure a system lifetime of several weeks, the device must be battery-operated, exhibit a compact form factor, and have low total system power dissipation.

Contrary to most existing systems, we seek high single-word detection accuracy, while striving to keep the energy footprint of each patient device to a minimum. In this demonstration proposal, we outline hardware/software architecture for single word speech recognition on an embedded device exhibiting severe processing and memory constraints. Despite these resource constraints, our prototype achieves a high recognition rate of 96%, while only dissipating an average of 28.5mW for continuous hardware based audio sampling and duty-cycled software-based speech recognition. Finally, we describe in detail how we plan to demonstrate our prototype at IPSN.

II. LITERATURE SURVEY

The design and implementation of prototype hardware/ software architecture for automatic single word speech recognition on resource constrained embedded devices. Designed as a voice-activated extension of an existing wireless nurse call system, our prototype device continually listens for a pre-recorded keyword, and uses speech recognition techniques to trigger an alert upon detecting a match. Here in order to demonstrate it will use advanced Linux environment and control the system with speech. We will have a detail description of the existing and the proposed methods to enable the extension applied in the real time implementation of the project.

A. Existing Method

In the existing system, they have not provided any algorithm or implementation procedure to execute the concept. The implementation is based on research of hardware system execution of output without any interface of software in the system. They have used hardware codec system in order to decode the voice stream which will eliminate in our proposed one.

Hardware: Fig.1 illustrates the existing prototype, which features an ARM 32-bit Cortex-M4 microcontroller. Audio signal acquisition is performed using a dedicated audio codec connected to a microphone. The analog signal from the microphone is sampled at a rate of 8 kHz with 16-bit resolution. When the pre-recorded keyword is detected by the speech processing chain, an attached communication node sends an alert via LWB to the graphical interface just as if a button had been pressed.

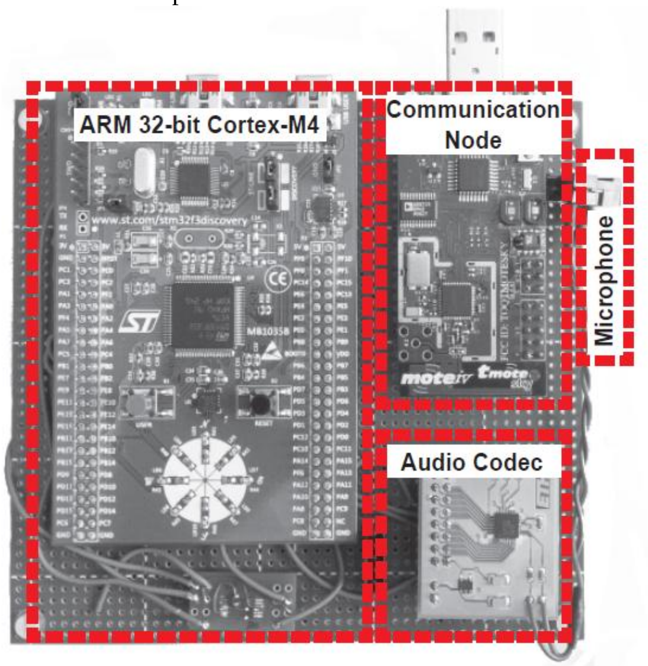

Fig.1.Prototype speech recognition device.

Software: In order to achieve the existing application requirements, the implementation is a state-of-the-art speech recognition algorithm, as shown in Fig.2, coupled with several power management techniques.

Fig.2. Speech Recognition processing chain.

Voice activity detection (VAD) uses an adaptive signal energy threshold to separate speech from background noise. When voice activity is detected, features of the digital audio signal are extracted. We use Mel-frequency cepstral coefficients (MFCCs) as the set of features. The recognizer compares these features to the stored keyword features and decides if the keyword had been spoken or not. Using dynamic time warping (DTW), we account for the keyword being spoken at different speeds. In order to reduce the power dissipation, we use direct memory access (DMA) transfers for storing the acquired audio signals into memory. As a result, the speech processing chain only needs to execute once every 128 ms, thus allowing aggressive duty cycling of the microcontroller. Additionally, we use dynamic frequency scaling to further reduce the system's power dissipation.

Performance: The two key performance metrics in our application scenario are power dissipation and keyword detection rate. Extensive measurements show that our prototype, without wireless communications, achieves continuous signal processing with a microcontroller duty cycle of 4.6% and an average system power dissipation of 28.5mW.

To evaluate the keyword detection rate, we perform a 5-fold cross validation using audio samples from five different speakers and four different languages. In order to mimic the target environment (i.e., a night-time dormitory room), we performed the experiments in a silent room. Each speaker chose a single keyword, which was recorded 20 times. A further 30 non-keywords were recorded once by each speaker. The audio samples were transferred from the prototype hardware into Matlab, where cross validation was performed. We find that our prototype achieves an average detection rate of 96%, while limiting the false accept rate to 1%. In summary, our system achieves higher detection accuracy and a lower energy footprint than existing embedded speech recognition systems.

Demonstration Details: Hardware/software prototype of a voice-activated wireless nurse call system and demonstrate its ability to detect a keyword, while giving insights into the speech processing pipeline. The demonstration exhibits two main components:

Prototype Device: We will showcase our prototype speech recognition device. Users will be able to configure it with a keyword of their choice and test the speech recognition by speaking the chosen keyword.

Graphical Visualization: We will extract debug information from the prototype in real-time and provide a graphical visualization of the captured audio samples, the voice activity detection, the extracted feature set, and the decision made by the speech recognition algorithm. Users will be able to observe the displayed information in real-time as well as pause the visualization for detailed inspection.

B. Proposed Methods

In proposed System, will execute the concept on ARM 11 Embedded device using Open Source Speech recognize engine. Also with Voice Input to user, the system will also respond to us accordingly. Two single words will be trained for the system to be identified when the user spokes on, i.e. Light, Fan. Whenever the user says the term the system will recognize the speech and turn on/off the device as per its status. To develop the software prototype of a voice-activated call system and demonstrate its ability to detect a keyword, while giving insights into the speech processing pipeline. The design and implementation of prototype hardware/software architecture for automatic single word speech recognition on resource constrained embedded devices. Designed as a voice-activated extension of an existing wireless nurse call system, our prototype device continually listens for a pre-recorded keyword, and uses speech recognition techniques to trigger an alert upon detecting a match. Here in order to demonstrate it will use advanced Linux environment and control the system with speech.

III. BLOCK DIAGRAM AND MODULES DESCRIPTION

Block diagram is as shown in Fig.3.

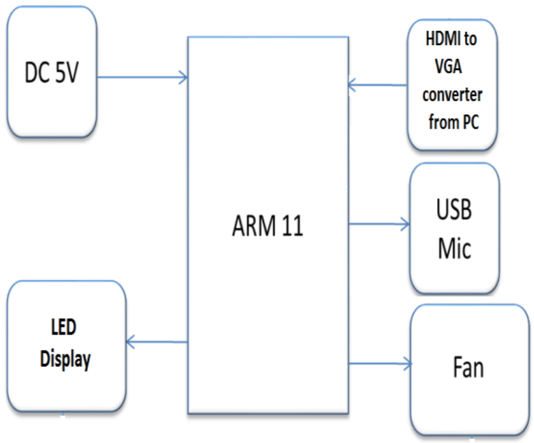

Fig.3. Block diagram of hardware architecture.

A. Power Supply

The input to the circuit is applied from the regulated power supply. The a.c. input i.e., 230V from the mains supply is step down by the transformer to 12V and is fed to a rectifier. The output obtained from the rectifier is a pulsating d.c voltage. So in order to get a pure d.c voltage, the output voltage from the rectifier is fed to a filter to remove any a.c components present even after rectification. Now, this voltage is given to a voltage regulator to obtain a pure constant dc voltage as shown in Fig.4.

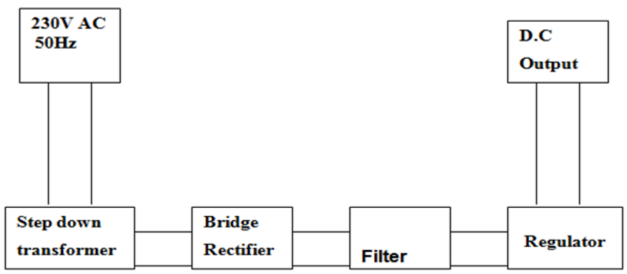

Fig.4. Power Supply.

B. Light Emitting Diodes

Light-emitting diodes are elements for light signalization in electronics. They are manufactured in different shapes, colors and sizes. For their low price, low consumption and simple use, they have almost completely pushed aside other light sources, bulbs at first place. They perform similar to common diodes with the difference that they emit light when current flows through them. It is important to know that each diode will be immediately destroyed unless its current is limited. This means that a conductor must be connected in parallel to a diode. In order to correctly determine the value of this conductor, it is necessary to know diode's voltage drop in forward direction, which depends on what material a diode is made of and what colour it is. As seen, there are three main types of LEDs. Standard ones get full brightness at current of 20mA. Low Current diodes get full brightness at ten time's lower current while Super Bright diodes produce more intensive light than Standard ones. Since the 8051 microcontrollers can provide only low input current and since their pins are configured as outputs when voltage level on them is equal to 0, direct connecting to LEDs is carried out as it is shown on fig.5 (Low current LED, cathode is connected to output pin). Microcontroller port pins cannot drive these LEDs as these require high currents to switch on. Thus the positive terminal of LED is directly connected to V_{cc}, power supply and the negative terminal is connected to port pin through a current limiting resistor. This current limiting resistor is connected to protect the port pins from sudden flow of high currents from the power supply.

Thus in order to glow the LED, first there should be a current flow through the LED. In order to have a current flow, a voltage difference should exist between the LED terminals. To ensure the voltage difference between the terminals and as the positive terminal of LED is connected to power supply V_{cc}, the negative terminal has to be connected to ground. Thus this ground value is provided by the microcontroller port pin.

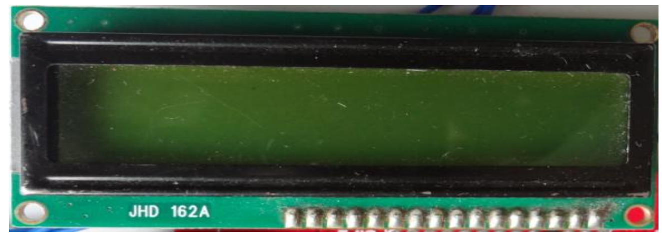

Fig.5. LCD.

C. ARM Processor

ARM (Acorn RISC machine) is a family of instruction set architectures designed for computer processors introduced by British company 'ARM Holdings' in 1985. ARM was developed in 1980s by British company Acorn computers which manufactured computers and was used in personal computers. BBC Micro series of computers were first to use ARM based coprocessor modules followed by Acorn computers and later dominated by IBM PC. In early days the available processors like Motorola 68000 and National semiconductor 32016 were unsuitable and lacking. Then Acorn inspired by Berkeley RISC project, decided to design a new architecture for its own processor. Acorn RISC machine project started in October 1983 using VLSI technology. First ARM silicon was produced on April 1985. The first ARM based computer 'Acorn Archimedes' was released in 1987 and Acorn won "Queen's Award for Technology" in 1992 for the ARM.

ARM Holdings develops architecture and Instruction set for ARM based products. Companies like Apple, Qualcomm, Nvidia, Samsung electronics and Texas Instruments make chips that implement ARM architecture. As of 2013, 10 billion ARM processors have been produced and 50 billion in 2014 representing 95% of smart phones, 35% of Digital Televisions and Set Top Boxes, and 10% of mobile computers. Based on architecture several cores have been designed by ARM holdings, they are as follows in Architecture - Core order. ARMv1- ARM1, ARMv2- ARM2, ARM3, ARMv3- ARM6, ARM7, ARMv4- ARM8, ARMv4T- ARM7TDMI, ARM9TDMI, ARMv5-ARM7EJ, ARM9E, ARM10E, ARMv6- ARM11 and so on Latest architecture developed in 2011 was ARMv8-A with 64/32 bit width and ARM Cortex-A53, ARM Cortex A57 are the cores designed by using this architecture. In our project we use ARM7TDMI core which is 32-bit width.

ARM7 Family: ARM7TDMI, ARM7TDMI-S, ARM720T, and ARM7EJ-S processors come under ARM7 family. The ARM7TDMI core is the most widely used 32-bit embedded RISC microprocessor. Optimized cost and power of ARM7TDMI makes most of the applications reliable and effective with low power consumption, small size, and high performance.

The ARM7EJ-S processor is a synthesizable core that provides all the features of the ARM7TDMI – small size, optimized power and the thumb instruction set including latest DSP extensions of ARM and enables acceleration of java-based applications. Strong-Arm® architecture software written for the ARM7TDMI processor core is 100% binary-compatible with all members of the ARM7 family. ARM7EJ-S is Forward-compatible with the ARM9, ARM9E and ARM10 families along with products of Intel's Strong ARM and x scale architectures. Thus designers can choose any of the available software-compatible processors as per the requirement of the application accounting price-performance. ARM architecture is supported by the following:

- Windows CE, Linux, palm OS and SYMBIAN OS and other operating systems.
- Above 40 RTOS, including Wind River's vxworks, qnx and mentor graphics' vrtx etc.
- Eda vendors which provide co-simulation tools.
- Different kinds of software development tools.

ARM7TDMI:

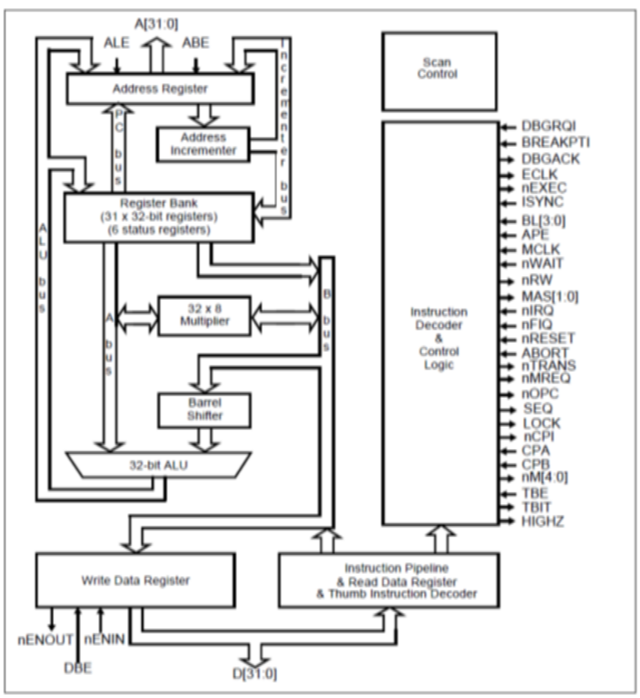

Fig.6. ARM7TDMI Core Diagram.

Fig.6 shows the ARM7TDMI Core Diagram. It is based on the Von- Neumann architecture with both instructions and data carried by Data bus. Data from memory is accessed by Load, store, and swap instructions. 8-bit, 16-bit, and 32-bit data is carried.

LPC2148 Microcontroller: LPC2148 microcontroller board is based on a 16-bit/32-bit ARM7TDMI-S CPU with real-time emulation and embedded trace support, that combine microcontrollers with embedded high-speed flash memory ranging from 32 KB to 512 KB. A 128-bit wide memory interface and unique accelerator architecture enable 32-bit code execution at the maximum clock rate as shown in Fig.7. For critical code size applications, the alternative 16-bit Thumb mode reduces code by more than 30% with minimal performance penalty. The meaning of LPC is Low Power Low Cost microcontroller. This is 32 bit microcontroller manufactured by Philips semiconductors (NXP). Due to their tiny size and low power consumption, LPC2148 is ideal for applications where miniaturization is a key requirement, such as access control and point-of-sale.

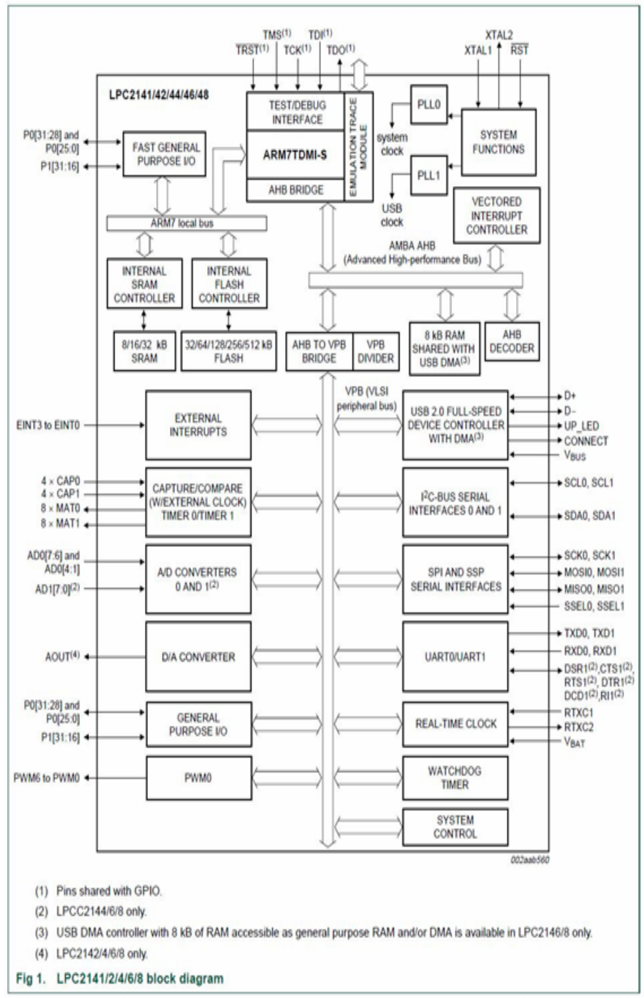

Fig.7. Block Diagram.

Features of LPC2148 Microcontroller:

- 16-bit/32-bit ARM7TDMI-S microcontroller in a tiny LQFP64 package.
- 8 KB to 40 KB of on-chip static RAM and 32 KB to 512 KB of on-chip flash memory; 128-bit wide interface/accelerator enables high-speed 60 MHz operation.
- USB 2.0 Full-speed compliant device controller with 2 KB of endpoint RAM. In addition, the LPC2148 provides 8 KB of on-chip RAM accessible to USB by DMA.

- One or two (LPC2141/42 Vs, LPC2144/46/48) 10-bit ADCs provide a total of 6/14 analog inputs, with conversion times as low as 2.44 ms per channel.
- Single 10-bit DAC provides variable analog output (LPC2148 only).
- Two 32-bit timers/external event counters (with four capture and four compare channels each), PWM unit (six outputs) and watchdog.
- Low power Real-Time Clock (RTC) with independent power and 32 kHz clock input.

IV. IMPLEMENTATION & RESULTS

The implementation and its results are explained in the form of steps as follows:

Step 1: The connections of the project are as displayed

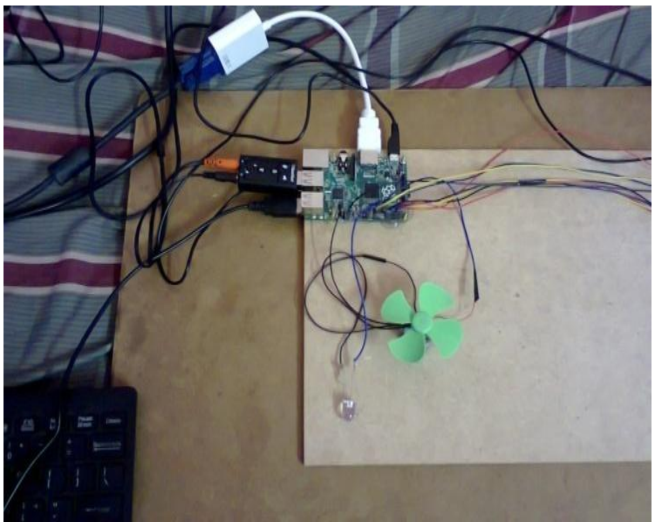

Fig.8.

Step 2: The kit is connected with the PC and the power supply is ON. Then the Raspbian OS is loaded.

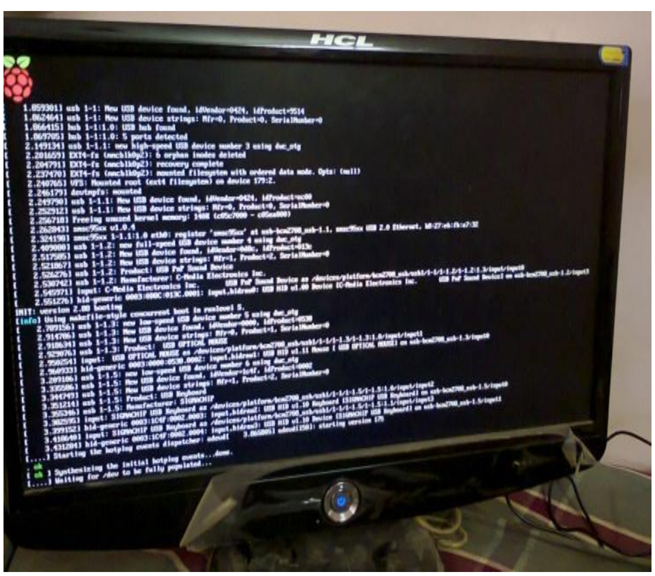

Fig.9.

Step 3: After loading the configuration, it displays the files configured with the OS.

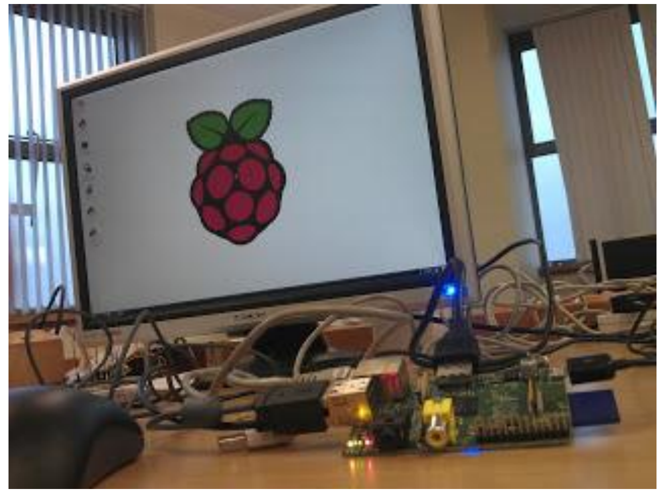

Fig.10.

Step 4: Open the voice recognition document in the project and follows the steps.

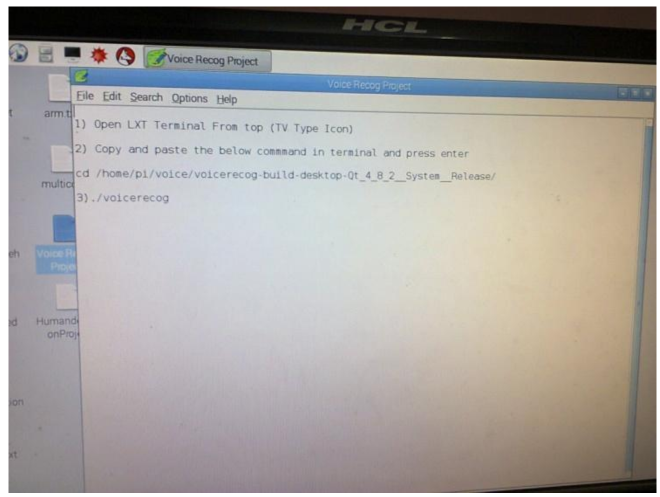

Fig.11.

Step 5: Follow the steps in the terminal by executing.

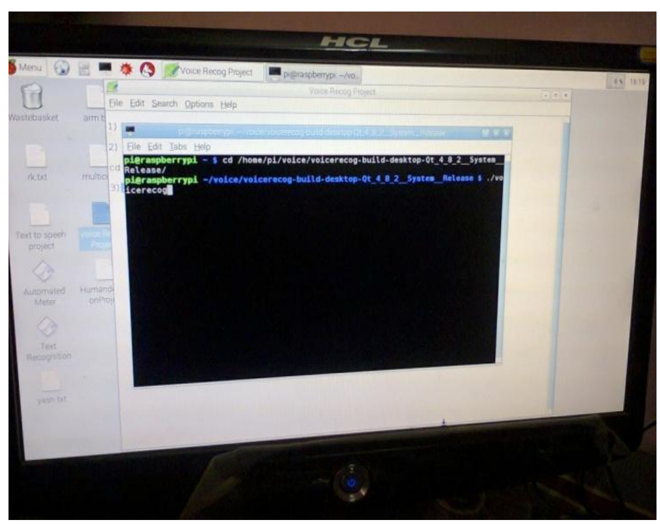

Fig.12.

Step 6: By executing the command in the terminal, we will find loading of terminal with main window open

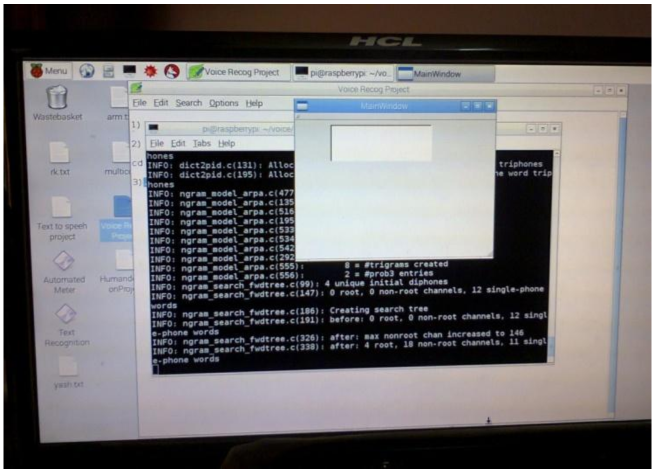

Fig.13.

Step 7: When we say Light On in stereo headphone, it displays on Main window and indicates speech recognition simultaneously by displaying LED as the output.

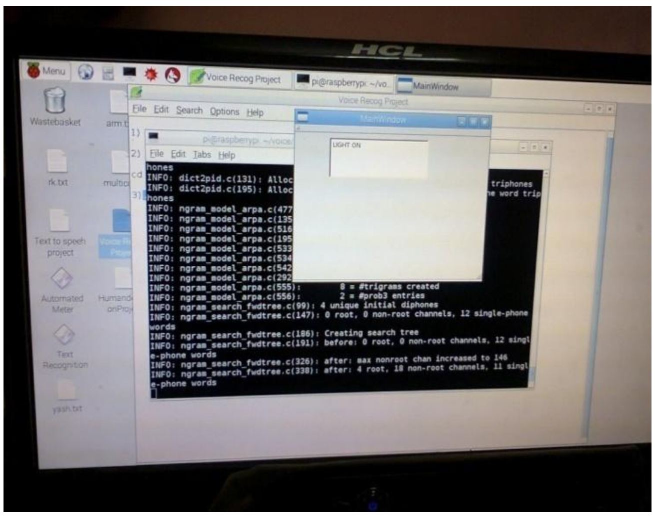

Fig.14.

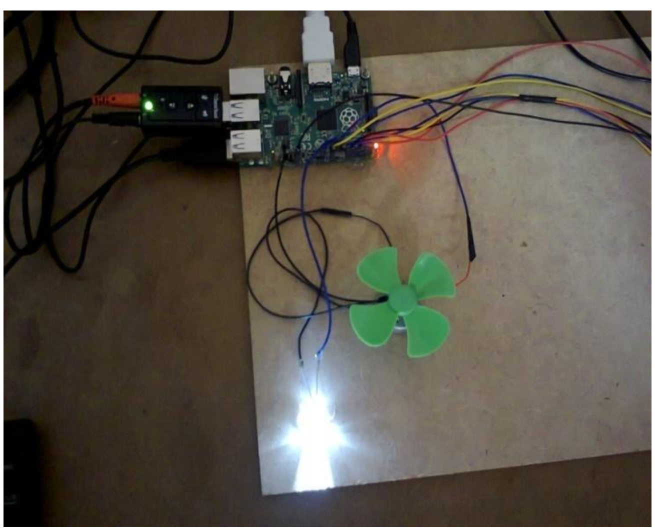

Fig.15.

Step 8: When we say Light OFF in stereo headphone, it displays on Main window and indicates speech recognition simultaneously by LED getting OFF in the output.

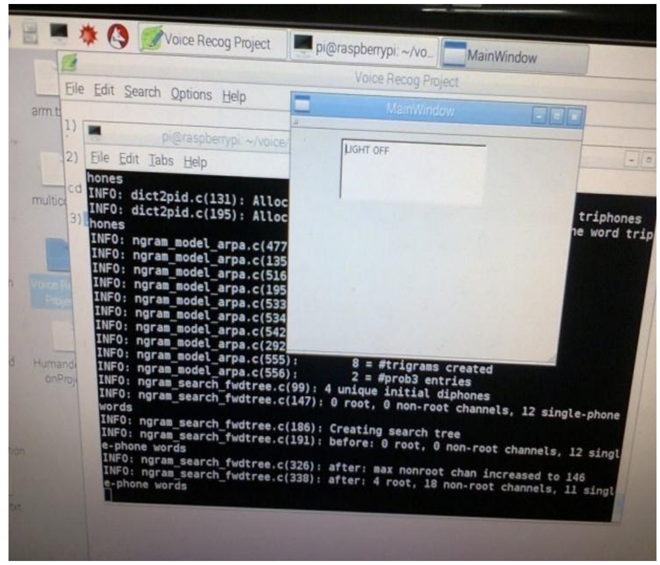

Fig.16.

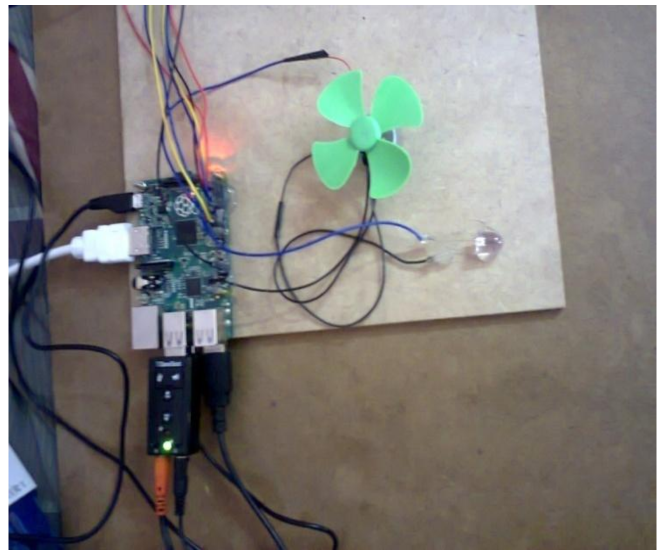

Fig.17.

Step 9: When we say Fan ON in stereo headphone, it displays on Main window and indicates speech recognition simultaneously by rotating fan in the output.

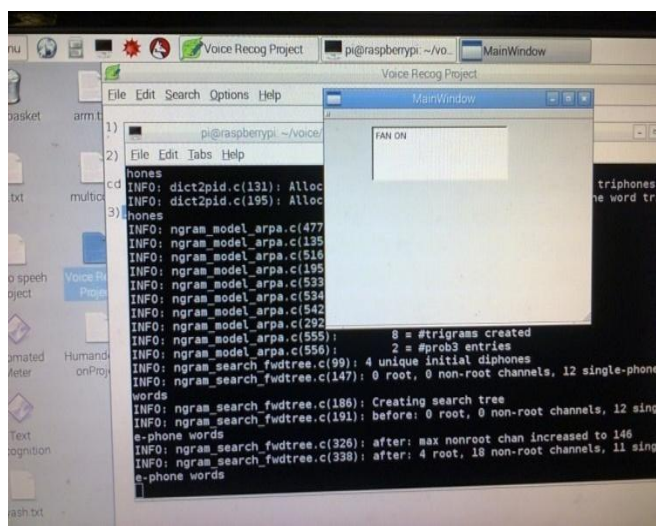

Fig.18.

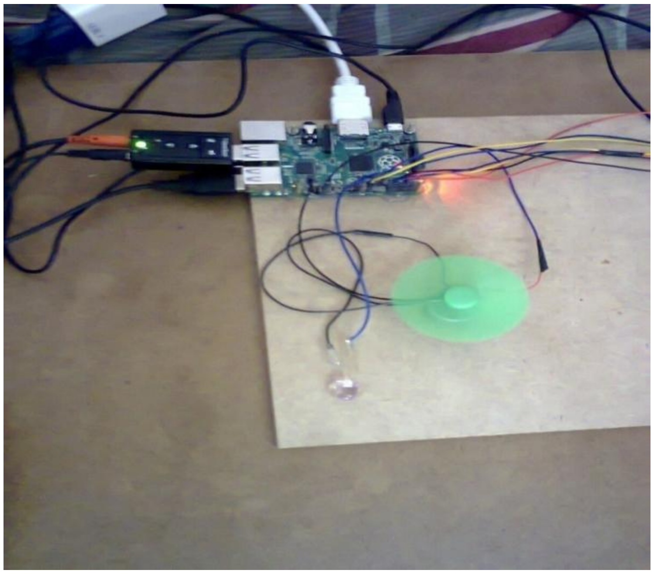

Fig.19.

Step 10: When we say Fan OFF in stereo headphone, it displays on Main window and indicates speech recognition simultaneously by stopping the fan rotating in the output.

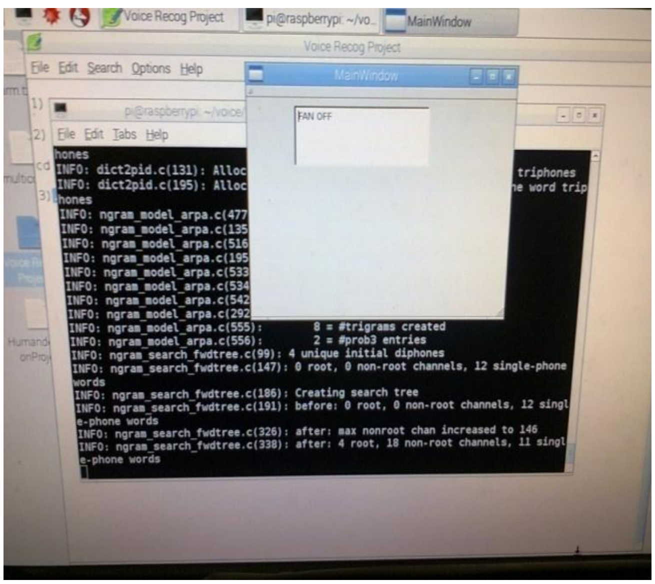

Fig.20.

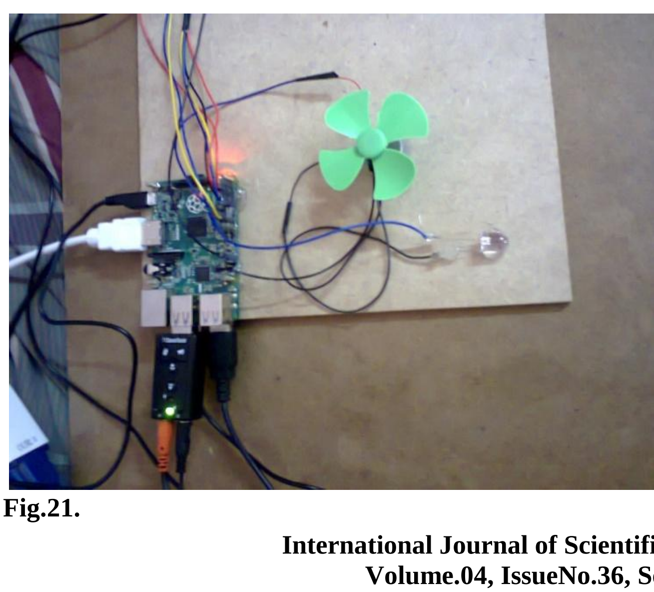

Fig.21.

V. CONCLUSION AND FUTURE SCOPE

The implementation of prototype hardware/software architecture for automatic single word speech recognition on resource-constrained embedded devices the objective of this project was mainly to build a speech recognizer for reception of particular word by the hardware and implementation of the task. In order to meet this objective, the data is stored in the Raspbian OS of the hardware kit Raspberry which is executed with the relevant data stored. The data stored in this project is the glowing of LED on & off and rotating the fan on & off through speech recognition. This is implemented in the project which can be used in wide variety of real time applications.

Future Scope: The future Scope implementation will be in wireless module for long distance with accurate sensors. The accuracy for detection of the words could be higher even through large distance execution of output which makes things simpler in case of multiple operations. This could also be implemented by recognizing various different languages and make the operations to perform in an accurate manner. Various languages data can be stored in database by creating a dictionary and multiple operations can be performed by different region people which can be provided to the customer satisfaction of requirement.

VI. REFERENCES

[1] J. Koet al., “Wireless sensor networks for healthcare,” Proc. IEEE, vol. 98, no. 11, 2010.

[2] M. Zimmerlinget al., “A reliable wireless nurse call system: Overview and pilot results from a summer camp for teenagers with Duchene muscular dystrophy,” in SenSys, 2013.

[3] F. Ferrari, M. Zimmerling, L. Mottola, and L. Thiele, “Low-power wireless bus,” in SenSys, 2012.

[4] O. Cheng, W. Abdulla, and Z. Salcic, “Hardware-software code sign of automatic speech recognition system for embedded real-time applications,” IEEE Trans. Ind. Electron., vol. 58, no. 3, 2011.

[5] D. Jurafsky and J. H. Martin, Speech and Language Processing. Prentice Hall, 2008.

[6] Deshmukh, N., Ganapathiraju, A, Picone J., (1999), Hierarchical Search for Large Vocabulary Conversational Speech Recognition. IEEE Signal Processing Magazine, 1(5):84-107.

[7] Dix, A.J., Finlay.J, Abowd, G., Beale, R. (1998). Human-Computer Interaction, 2nd edition, Prentice Hall, Englewood Cliffs, NJ, USA.

[8] Kagaba,S., Nsanzabaganwa, S., Mpyisi,E., (2003), Rwanda Country Position Paper, Regional Workshop on Ageing and Poverty Dares Salaam, Tanzania.

[9] Kandasamy, S., (1995), Speech recognition systems. SURPRISE Journal, 1(1).

[10] Liu, F.H., Liang G., Yuqing G. AND Picheny, M, (2004). Applications of Language Modeling in Speech-To-Speech Translation International Journal of Speech Technology (7):221-229.

Author's Details:

Rajulapati Poorna Raghava Rao, Obtained B.Tech Degree in Electronics And Communication Engineering From D.M.S.S.V.H College Of Engineering, Machilipatnam And pursuing his M.Tech From CMR Technical Campus, Hyderabad. His Area of Interest Embedded system design, Embedded C, Java programming language.

Kesana Mohana Lakshmi M.TECH, (Ph.D) Working As A Associate Professor for Department of Electronics And Communication And Engineering in CMR Technical Campus, Hyderabad And doing his Ph.D Work in Acharya Nagarjuna University (ANU) Obtained B.Tech Degree in Electronics And Communication Engineering From Gokaraju Rangaraju Institute of Engineering & Technology.Recived his M.tech JNTU Hyderabad (JNTUH) Having 6 years of experience in Digital Image Processing, Embedded Systems.