Development of a stand-alone speech recognition SoC to assist in mail and courier sorting

Tanvi Sahay, *Arpit Aggarwal
Mentor: Dr. Mahesh Chandra
Birla Institute of Technology Mesra, Ranchi – 835215
***Email: ar13pit@gmail.com**

***Abstract*—**The present system of manual sorting of packages is the cause of some of the most common reasons for delay in delivery of mails and courier packages. These include unavailability of complete and legible addresses on the packages and misplacement of packages at the sorting facility itself, which is a result of the unavoidable factor which comes into play at each level i.e. the factor of human error. The aim of this project is to develop a reusable voice recognition system on chip for the purpose of sorting of packages in the central and local sorting hubs of logistics companies, to enhance their productivity and minimize the chances of delay in delivery. The project employs a speech storage and recognition chip to store a consignee's address at the packaging facility, which is then attached to the corresponding packet. This chip will act as a beacon to let the employees at the sorting facility locate a packet by simply calling out its destination address, thus minimizing the possibility of misplacement.

Keywords: Speech Recognition, System on chip, HM2007, sorting

I. INTRODUCTION

Speech recognition is one of the most sought after technologies in the world today. From voice activated home automation systems [1], voice activated automated dialling to voice controlled GPS systems, speech, by virtue of being one of the most fundamental characteristics of homo-sapiens has been convolved with science to present us with innovative solutions to some of the most trivial problems. The techniques most prevalent for speech processing include the use of mathematical models such as Hidden Markov models (HMM), Gaussian Mixture Models (GMM) and Artificial Neural Networks (ANN) to recognize patterns in the voice of different individuals and use these recognized patterns to perform certain tasks. All the aforementioned techniques have been successfully applied for development of speech processing systems and have been used to create various applications and software that aid us in our daily lives. But in contrast to all the advancement going on the software front, hardware development represents only a mere fraction of speech processing research. Although some unique devices such as the Talking Toaster [3] and voice activated wheelchair [2] have been developed, fabrication of stand-alone systems still remains a largely untreated territory.

This project aims at development of such a stand-alone System-On-Chip around the ANN based voice recognition chip, HM2007 [1-6] which will help solve the problem of sorting in central as well as local hubs of all the major logistics companies.

The present system of manual sorting of packages is prone to faults such as erroneous, incomplete or illegible addresses and an occasional shortage of labour which cause delay in the delivery of packages thus troubling the consignor, consignee as well as the service providers. At times, human error results in placement of a package in the wrong set of packages, which causes further delay in the delivery of the package.

This chip aims at bypassing the above mentioned problems to maximize productivity of each centre and thus increase its capacity to handle mails. It will not just ease the sorting process but also make it possible to search a particular packet from a heap of packets which otherwise is a very cumbersome and time-taking job.

Design objectives for this project to be applicable for real-time working include:

- Optimized size to avoid hampering with the courier/mail package
- Reusable chip to minimize investment cost for the companies
- Customizable for different courier agencies
- Provision for modification to employ it in other sorting and locating facilities.
- Ease of operation to make it suitable for personal use.

II. TECHNICAL BACKGROUND

Sorting of packages is largely a two-tier process. The first process involves sorting according to different countries in case of international transactions and according to different states in case of national transactions. The next stage is sorting within the local domain i.e. within the country or within the state. Once the packages reach the state level offices, manual sorting is done to separate the packets pertaining to a particular city or district which are then sent to their respective local offices for further sorting.

The national as well as international facilities are largely mechanized and sorting is an automated process, being realized with the help of machines which scan bar codes on each package. A message from this reader goes to the next stage of processing where mechanized arms place the package into the section reserved for all packages of this particular area. A bundle of these packages is then formed and loaded onto trucks which carry these to the local offices to be further sorted accordingly. Packages from the international hubs are first shipped out or transported through air to their respective countries. The packages are then separated according to states, followed by city wise sorting at each state level facility.

Each city has a central sorting facility and a number of local sorting facilities. Sorting in both these facilities is done manually i.e. labour is employed to manually read the addresses on each packet and sort them accordingly in what are known as ‘pigeon hole racks’. A typical central hub employs 5-15 workers who work in shifts to sort the consignments that come in at any time of the day. These sorted packets are taken to the local distribution centre from where they are collected to be hand delivered by the delivery officers.

III. PROPOSED SOLUTION

This project proposes to attach a voice recognition chip to each consignment at the main collection facility itself and store certain words in this chip, which will indicate the destination address of the package. The chip will be attached inside a special box in which each package will be inserted, before being transported to the next lower level hub. The external packaging will ensure that no harm is done to the consignment. Also, once the package has been delivered, this box can be collected back by the delivery officer. These collected boxes can then be distributed to the different post offices and courier service centres within the city where they can be reprogrammed to store a different address and can hence be reused over and over again.

This method presents several benefits over the traditional method of sorting.

- The packets will only need to be called out, in response to which they will automatically indicate their location, making it much easier to separate packets belonging to a certain domain, thus saving a crucial amount of time.
- At present, a delivery officer needs to sort packages according to his route before he actually takes to deliver them. With this chip, the officer can skip this step as all he would now need to do is speak the name of the consignee

and the parcel will automatically indicate its location.

- This project will enhance the process of time based deliveries, a factor on which a lot of franchisees are rated.
- The project presents a one-time investment cost owing to the reusable feature of the chip, thus reducing the investment of a service provider.
- This is an easy to adopt method as it does not interfere with any of the current hub selection and transportation networks

IV. IMPLEMENTATION

The project was successfully implemented by dividing the entire working circuit into multiple parts and realizing each one separately. The final circuit, made by the integration of all the sub-circuits, consist of the following parts:

A. The indicator circuit

This circuit consists of a buzzer and multiple LEDs interfaced with a TIVA C series Development Board to work as an indicator when a word pertaining to the stored address is recognized. Fig. 1 shows the final design of this circuit implemented on a printed circuit board.

B. The voice input circuit

A condenser microphone has been used to take voice input from the user. Output of the microphone, after appropriate conditioning, has been given to an inverting amplifier with a gain of 47 to make the signal detectable by HM2007. Fig. 2 shows the working configuration of this microphone.

C. The speech recognition circuit

The central unit of this circuit is the voice recognition chip HM2007. This chip has a capacity to store 20 isolated words of 1.92 seconds or 40 isolated words of 0.96 seconds each. The chip works on the principle of Artificial Neural Networks, although an exact configuration of this IC is yet unknown. This circuit was again implemented in two stages, the first being implementation of manual mode of HM2007, controlled with the help of switches or buttons. A detailed description of the working of this chip was formulated through experimentation in this mode. The second stage was controlling the chip and its recognition through the TIVA C series Development Board, which has been explained later.

The voice recognition circuit, as shown in fig. 3, consists of HM2007 along with a 64K SRAM which stores the words to be recognized. A brief description of working of this IC has been given below:

The IC works on two basic commands, CLEAR and TRAIN.

- CLEAR command is used to clear contents of a particular memory location or of all the locations, depending on what the user wishes to do.
- TRAIN command is used to train the chip with a particular word by first providing it with a memory location at which it will store the word and then speaking the word to be stored.

This chip is by default in recognition mode i.e. it keeps taking voice inputs and keeps matching them with pre-stored values to recognize words. It has an acknowledged system i.e. acknowledgement is provided to the user for each voice input taken by the chip. This acknowledgement can either be memory location of the recognized word or one of the three error codes, denoting an unrecognized word or notifying the user when the word spoken by them is either too long or too short.

Efficiency of correct recognition of words is directly proportional to the number of words for which HM2007 has been trained. The reason for this is that HM2007 not only recognizes the word but also provides a ‘score’ of recognition which denotes the extent to which a word matches with another one stored in the memory. Thus the chip may claim to recognize a word that has not been stored in the memory but in a case as that, the score of recognition would be very low. It is thus advisable to store more than 7 words for good recognition efficiency.

Fig. 4 shows the final PCB, complete with a transistor network to take inputs from MCU and a latch to give inputs to it.

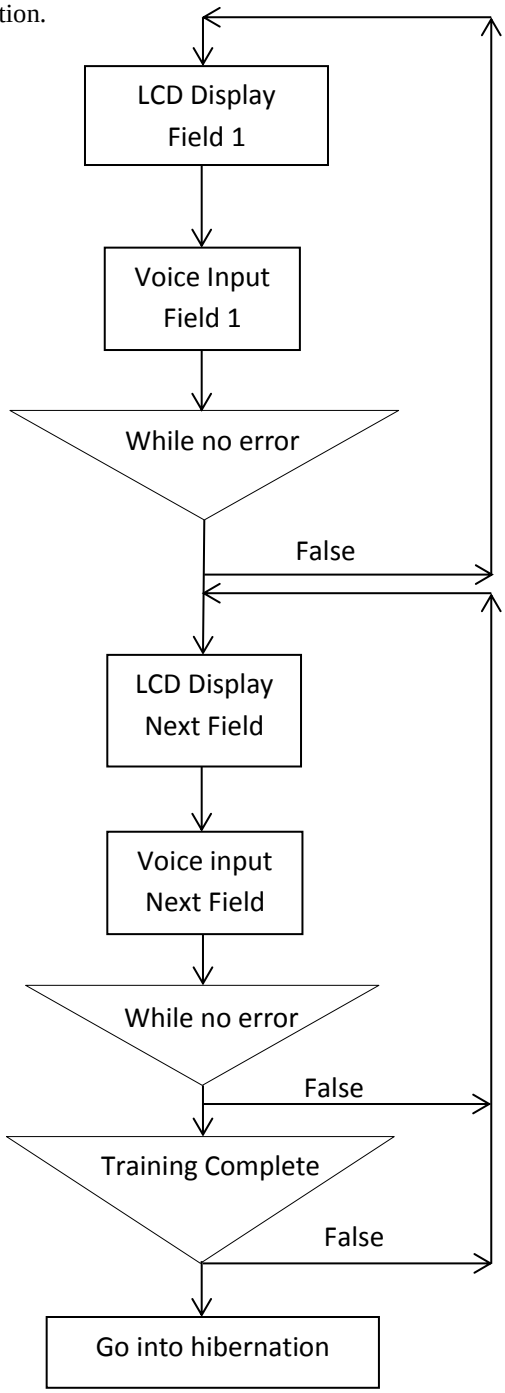
D. The Stand-alone System

This is the final piece of the system, integrating all other parts together and controlling them through the tm4c123gh6pge MCU used in the development board. This board controls the entire working of the circuit and switches the buzzer and LEDs ON upon recognition of the word. Fig. 6 shows the final circuit, complete with the connection of HM2007 and indicator circuit with TIVA Development Board. Outputs to the board are given via a latch to match the logic levels of the IC and the Development Board. Commands are given to drive the entire system through this Board, which acts as the master control unit. The on-board LCD is used to display which field is to be stored by the speaker. Once the field is displayed, sufficient delay is provided for the user to be able to read the words and prepare him/ herself for entering that particular word. Blinking of an LED will indicate when the speaker is supposed to enter that word, thus avoiding any overlap between words which might confuse the system. The fields to be stored in the system have been pre-defined as:

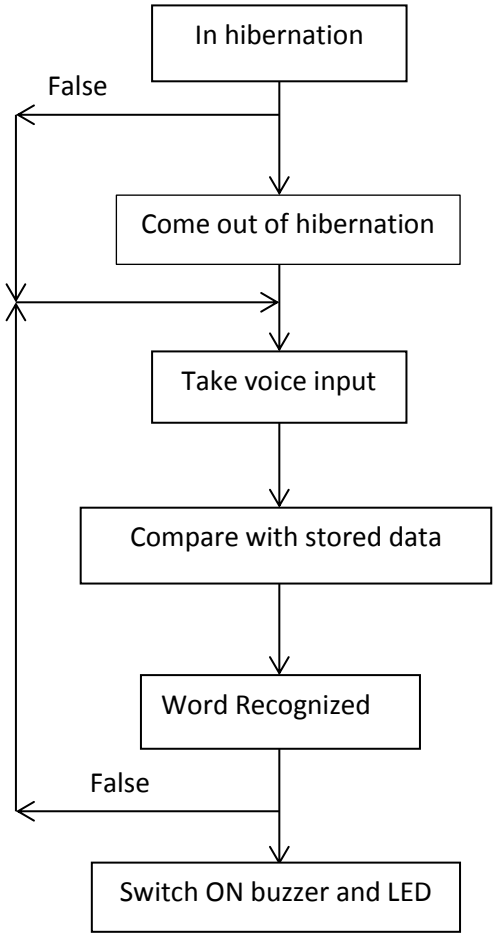
1. Last Name 1 field
2. First Name 1 field
3. Street Name 2 fields
4. Area Name 2 fields
5. City 1 field
6. State 1 field

As one can see, these fields do not incorporate the entire address of a consignee. This is due to the fact that this chip is being used for assistance of the service providers and not the consumers, hence only fields important for sorting and locating packages have been used. Since a delivery officer has a fixed route, he already knows which houses to stop at in which case house number would be an unnecessary field. Pin-code has also been omitted from the above entries as in many cases, multiple areas have the same pin-code, which lead to confusion. Also, sorting of packages on the lower levels, for where this system has been designed, is done according to areas, not pin-codes which thus makes it a useless field as well.

Once all words have been stored, the development board is put into hibernate mode, so that the voice recognition circuit is inactive during the transit of packages. A voice activated trigger pulls the board out of the hibernation mode once it reaches its destination.



Flowchart for training mode



Flowchart for recognition mode

V. RESULTS

The system was tested for 10 distinct addresses and a recognition efficiency of **92.6** percent was obtained. Following table shows the results for efficiency of recognition for different input fields. This system works with a good efficiency for moderately noisy signals as well as low noise signals. The only drawback to the present system is that its behaviour is unpredictable in case of a high noise environment.

INPUT FIELD	AVERAGE EFFICIENCY
Last name	91.2%
First name	92.4%
Street name line 1	95%
Street name line 2	92.8%
Area name line 1	90.7%
Area name line 2	91.4%
City	90%
State	97.3%

VI. CONCLUSION

From the above results and numerous other experiments, it can be concluded that the stand-alone voice recognition chip was successfully designed and can be used for the sorting of packages in real time, provided the level of noise is within its bounds. The system has the capability to reduce time wasted in manual sorting. As processing time of the chip is very low, instant output is available at the buzzer and LED without any time wasted in unnecessarily reading each package .

VII. REFERENCES

[1] V. Ramya, B. Palaniappan,"Embedded Home Automation for Visually Impaired", IJCA, Volume 41 – No. 18, March 2012.

[2] "Voice Activated Multiprocessor Embedded System to Improvise The Control of Motorized Wheel Chair", Sangmeshwar S. Kendre et. al./International Journal of Engineering Science and Technology, Vol.2(11),2010,6812-6818

[3] Chris Setter, Corin Anderson. Project Talking Toaster – Final Report, CSE 447, Spring 1996, University of Washington.
http://www.the4cs.com/~corin/cse477/toaster/

[4] Kon-Hyong Kim, Tae Yong "Tim" Chung, Kevin Jin-Ho Ham. Project Checkmate, ECE 476, Spring 2007, Cornell University.
https://courses.cit.cornell.edu/ee476/FinalProjects/s2007/tc228_kk336_kh272/tc228_kk336_kh272/Hardware.html

[5] Justin Thomas Wodark,"A HARDWARE IMPLEMENTATION OF AN ARTIFICIAL NEURAL NETWORK",CaliforniaState University, Northridge
[6] Dinu Mathew," Voice Controlled Data Acquisition Car Based on ZIGBEE Technology", IOSRJECE, ISSN: 2278-2834 Volume 2, Issue 3 (July-Aug 2012), PP 19-24
[7] HM2007 Datasheet
[8] HY6264A-LP 70 Datasheet
[9] SN74F373N Datasheet

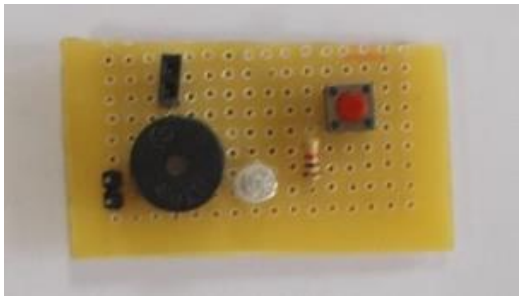


Fig. 1 The Indicator Circuit

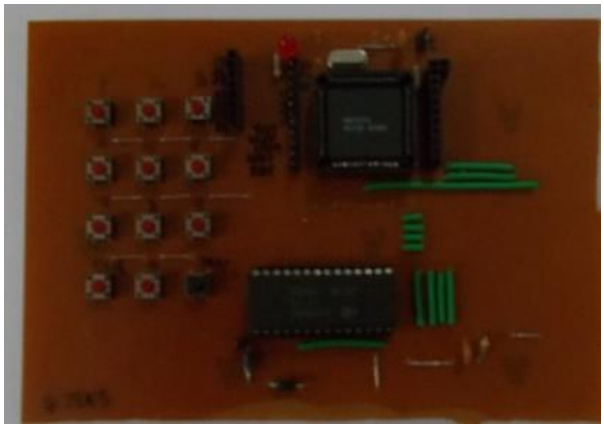


Fig. 2 The voice recognition circuit

Fig. 3 The Speech Recognition Circuit

Fig. 4 The final PCB

BILL OF MATERIALS

Material	Cost (Dollars)
UA741CP	0.19
Condenser Microphone	0.16
BC547	0.032
7805 voltage regulator	0.17
tm4c123gh6pge microcontroller	150
HM2007	20.0
SN74F373N Latch	0.53
HY6264 64K SRAM	0.80
3v lithium battery	0.62
Capacitors and resistors (lump sum)	2.00
DC Jack	0.08
Crystal Oscillator	0.24
9v rechargeable battery	5.57
Miscellaneous Components	0.50