



Software Independent Speech recognition System

Pranali Yawle¹, Devika Pawar², Pooja Pawar², Puja Dhumal²

Assistant Professor, Dept. of E&Tc, BVCOEW, Pune, India¹

UG Student, Dept. of E&Tc, Bharati Vidyapeeth's College of Engineering For Women, Pune, India²

ABSTRACT: Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format or digital codes. Rudimentary speech recognition software has a limited vocabulary of words and phrases and may only identify these if they are spoken very clearly.

The speech recognition system is completely assembled and easy to use programmable speech recognition. Programmable, in the sense that you train the words (or vocal utterances) you want the circuit to recognize. This allows you to experiment with many facets of speech recognition technology. It has 8 bit data out which can be interfaced with any microcontroller for further development. Some of interfacing applications which can be made are controlling in-car systems, health care medical documentation, Speech Assisted technologies, Speech into text translation, High-performance fighter aircraft and many more [1].

KEYWORDS: Speech recognition, Microphone, Voice,

I. INTRODUCTION

Speech recognition applications include voice user interfaces such as voice dialling (e.g. "Call home"), call routing (e.g. "I would like to make a collect call"), domestic appliance control, simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input). Speech recognition is the translation of spoken words into text. It is also known as "automatic speech recognition", "ASR", "computer speech recognition", "speech to text", or just "STT". Speech Recognition is technology that can translate spoken words into text. Some SR systems use "training" where an individual speaker reads sections of text into the SR system. These systems analyse the person's specific voice and use it to fine tune the recognition of that person's speech, resulting in more accurate transcription. At present, there have been a number of successful commercial voice interfaces. The most prominent example is Siri, the voice-activated personal assistant built in the latest iPhone. Speech recognition products are also available in Android, the Windows Phone platform, and most other mobile systems with considerable limitations. The recognition accuracy and performance of a system would degrade dramatically with small modifications of speech signal or speaking environment. As a result more computation and memory capacity are needed for Speech recognition. Especially, it is useful for embedded systems like smart phones and PDAs having insufficient space for typing or touching and helpful for controlling navigation during car driving. Also, it can be used to build advanced security systems and ATM machines [2]

The heart of the circuit is the Hm2007 speech recognition IC. The IC can recognize 20 words, each word a length of 1.92 seconds. This system is based on Hm2007 IC.

II.BACKGROUND

Speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data. The advances are evidenced not only by the surge of academic papers published in the field, but more importantly by the world-wide.

Industry adoption of a variety of deep learning methods in designing and deploying speech recognition systems. These speech industry players include Microsoft, Google, IBM, Baidu (China), Apple, Amazon, Nuance, IflyTek

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

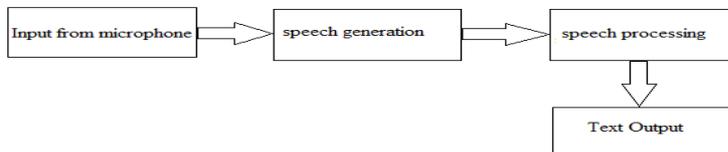
(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

(China), many of which have publicized the core technology in their speech recognition systems being based on Deep learning.

III. PROPOSED METHODOLOGY AND DISCUSSION

Speech is an exceptionally attractive modality for human computer interaction: it is “hands free”; it requires only modest hardware for acquisition (a high-quality microphone or microphones); and it arrives at a very modest bit rate

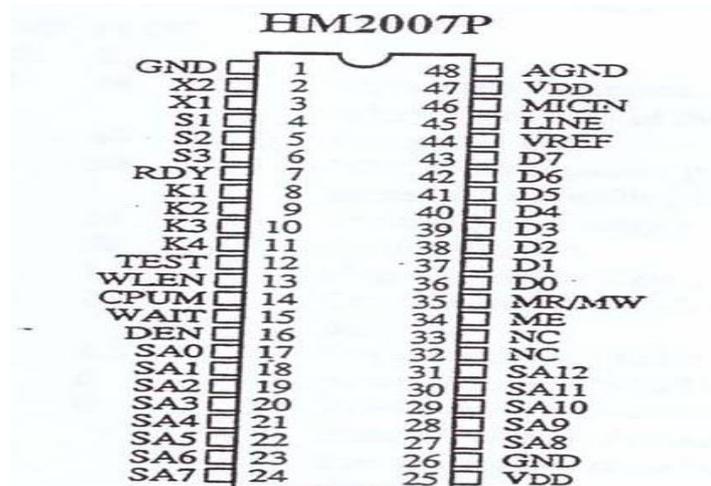


A).Fig. Basic block diagram of speech to text converter

It takes speech data from microphone and then it converts into text format by using speech processing. Recognizing human speech, especially continuous (connected) speech, without burdensome training (speaker-independent), for a vocabulary of sufficient complexity (60,000 words) is very hard. However, with modern processes, flow diagram, algorithms, and methods we can process speech signals easily and recognize the text which is talking by the talker. In this system, we are going to develop an on-line speech- to-text engine [4]. The system acquires speech at run time through a microphone and processes the sampled speech to identify the uttered text. The recognized text can be stored in a file. It can supplement other larger systems, giving users a different choice for data entry.

B).Description of HM2007

The HM2007 is a CMOS voice recognition LSI (Large Scale Integration) circuit. The chip contains an analog front end, voice analysis, regulation, and system control functions. The chip may be used in a stand-alone or CPU connected. Features[8]



- Single chip voice recognition CMOS LSI
- Speaker dependent.

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

- External RAM support.
- Maximum 40 words recognition(96 sec)
- Maximum word length 1.92 sec
- Microphone support.
- Manual and CPU modes available

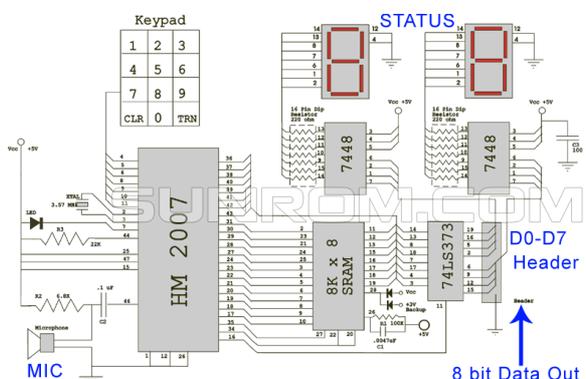


Fig. Schematic diagram of HM2007/ speech recognition

HM2007 is a single chip CMOS LSI circuit with on chip front end and analysis. Circuit is composed of using the external microphone, keyboard, 8x8KSRAM and basic components like seven segment display, latch etc. The schematic diagram of HM2007 speech recognition kit is described in above figure. [5]

Training of words:

Press “1” (display will show “01” and the LED will turn off) on the keypad, then press the TRAIN key (the LED will turn on) to place circuit in training mode, for word one. Say the target word into the headset microphone clearly. The circuit signals acceptance of the voice input by blinking the LED .off then on. The word (or utterance) is now identified as the “01” word. If the LED did not flash, start over by pressing “1” and then “TRAIN” key. You may continue training new words in the circuit. Press “2” then TRN to train the second word and so on. The circuit will accept and recognize up to 20 words (numbers 1 through 20). It is not necessary to train all word spaces. If you only require 10 target words that’s all you need to train .

A. Testing of words:

Repeat a trained word into the microphone. The number of the word should be displayed on the digital display. For instance, if the word “dictionary” was trained as word number 20, saying the word “dictionary” into the microphone will cause the number 20 to be displayed [1].

B. Error Codes:

The chip provides the following error codes.

- 55 = word to long
- 66 = word to short
- 77 = no match



International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

C. Recognition Style:

In addition to the speaker dependent/independent classification, speech recognition also contends with the style of speech it can recognize. They are three styles of speech: isolated, connected and continuous. Isolated: Words are spoken separately or isolated. This is the most common speech recognition system available today. The user must pause between each word or command spoken. Connected: This is a half-way point between isolated word and continuous speech recognition. It permits users to speak multiple words. The HM2007 can be set up to identify words or phrases 1.92 seconds in length. This reduces the word recognition dictionary number to 20. Continuous: This is the natural conversational speech we use to in everyday life. It is extremely difficult for a recognizer to sift through the sound as the words tend to merge together.

IV. EXPERIMENTAL RESULT

This sample project will show how a circuit can be interface through the data bus of speech recognition circuit. It will show message and error codes on LCD. It takes input from microphone in speech format convert into text by using IC HM2007.

V. CONCLUSION

This paper introduces the basics of speech recognition and speech to text technology and also highlights the difference between different speech recognition systems. In this paper the most common algorithms and basic block diagram which are used to do speech recognition are also discussed along with the current and its future use.

REFERENCES

- [1] Sunpreet Kaur,Nanda,Akshay,P.Dhande, "Microcontroller Implementation of a Voice Command Recognition System for Human Machine Interface in Embedded System", (IJECSCE) Volume 1, Issue 1
- [2] en.wikipedia.org/wiki/Speech recognition.
- [3] Miss.Prachi Khilari, "A review on speech to text conversion method", (IJARCET) Volume 4 Issue 7, July 2015
- [4] Santosh Gaikwad , Bharti Gawali , Suresh Mehrotra,, "Marathi Speech Interface System for the Activation and Controlling of Electronic Equipment", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue:
- [5] en.wikipedia.org/wiki/HM2007 datasheet
- [6] L. R. Rabiner, S. E. Levinson, A. E. Rosenberg and J. G. Wilpon, "Speaker Independent Recognition of Isolated Words Using Clustering Techniques", IEEE Trans. Acoustics, Speech and Signal Proc., Vol. Assp-27, pp. 336-349, Aug. 1979.
- [7] J. Suzuki and K. Nakata, "Recognition of Japanese Vowels— Preliminary to the Recognition of Speech", J. Radio Res. Lab, Vol. 37, No. 8, pp. 193-212, 1961.
- [8] M. Mohri, "Finite-State Transducers in Language and Speech Processing, Computational Linguistics", Vol. 23, No. 2, pp. 269- 312, 1997.
- [9] M.Marzinik and B.Kollmeir, "Speech Pause Detection For Noise Spectrum Estimation By Tracking Power Envelope Dynamics", IEEE Transactions On Speech And Audio Processing, Barcelona, Vol.10, No.2, Feb.2002, pp. 109-117.
- [10]en.wikipedia.org/wiki/Voice recognition using HM2007.

VII. KIT PHOTO

