

Speaker Independent Voice Recognition Calculator

Yahya S. H. Khraisat

AL-Balqa' Applied University,
Al-Huson University College,
Electrical and Electronics Department,
P.O. Box 50, Al-Huson 21510, Jordan
yahya@huson.edu.jo

Abstract

The voice activated calculator is a speaker-independent system that is used to perform basic mathematical operations. It recognizes the isolated spoken digits from 0 to 9, and other words like plus, minus, times, equal and clear. It then performs the respective arithmetic operations, and displays the final answer on an LCD display.

Keywords: PIC Microcontroller, Voice Recognition Calculator, LCD display and Speech detection and Recognition

1. Introduction

Speaker-independent voice recognition systems have a very strong probability of becoming a necessity in the workplace in the future. Such systems would be able to improve productivity and would be more convenient to use. The idea of a hardware that can recognize any person's voice without the training time involved in currently employed systems is a very promising one, and possibly a marketable one too.

At its most basic level speech recognition allows the user to perform parallel tasks, (i.e. hands and eyes are busy elsewhere) while continuing to work with the computer or appliance. Another aspect of this hardware would be in the assistance of hand-disabled people.

The block diagram of this real-time system is shown below in Figure 1.

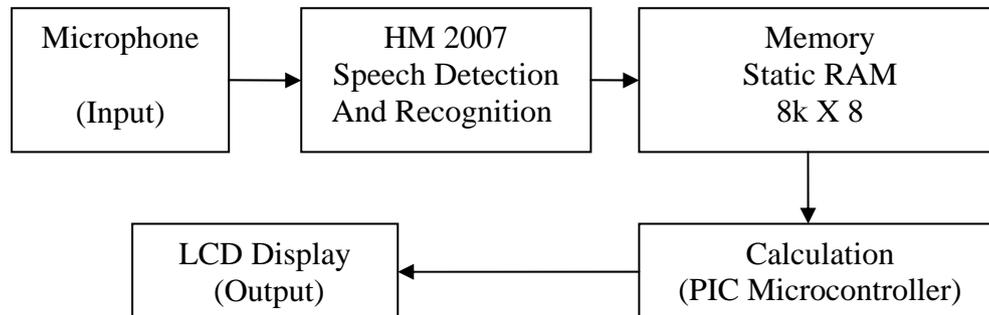


Figure 1 Block diagram of voice activated calculator

1.1 . Specifications

1.1.1. Speaker Dependent / Speaker Independent

Speech recognition is classified into two categories, speaker dependent and speaker independent.

Speaker dependent systems are trained by the individual who will be using the system. These systems are capable of achieving a high command count and better than 95% accuracy for word recognition. The drawback to this approach is that the system only responds accurately only to the individual who trained the system. This is the most common approach employed in software for personal computers.

Speaker independent is a system trained to respond to a word regardless of who speaks. Therefore the system must respond to a large variety of speech patterns, inflections and enunciation's of the target word. The command word count is usually lower than the speaker dependent however high accuracy can still be maintain within processing limits. Industrial requirements more often need speaker independent voice systems, such as the AT&T system used in the telephone systems [1].

Recognition Style

Speech recognition systems have another constraint concerning the style of speech they can recognize. There are three styles of speech: isolated, connected and continuous.

Isolated speech recognition systems can just handle words that are spoken separately. This is the most common speech recognition systems available today. The user must pause between each word or command spoken. The speech recognition circuit is set up to identify isolated words of 0.96 second lengths.

Connected is a half way point between isolated word and continuous speech recognition. Allow 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 vocabulary number to 20.

Continuous is the natural conversational speech we are used to in everyday life. It is extremely difficult for a recognizer to shift through the text as the words tend to merge together. For instance, "Hi, how are you doing?" sounds like "Hi,.howyadoin" Continuous speech recognition systems are on the market and are under continual development [2].

2. Testing and Validation

2.1 Training the Speech Recognition Circuit

The keypad and digital display are used to communicate with and program the HM2007 chip. The keypad is made up of 12 normally open momentary contact switches. When the circuit is turned on, "00" is on the digital display, the red LED (READY) is lit and the circuit waits for a command.

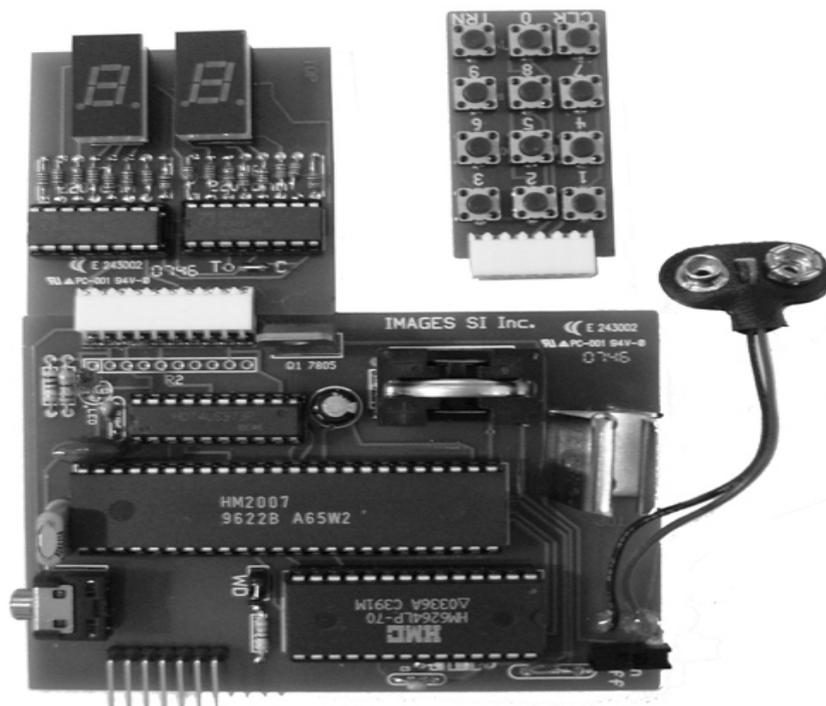


Figure 2 The Speech recognition System

The Complete Schematic of Speech Recognition Circuit is shown below in figure3.

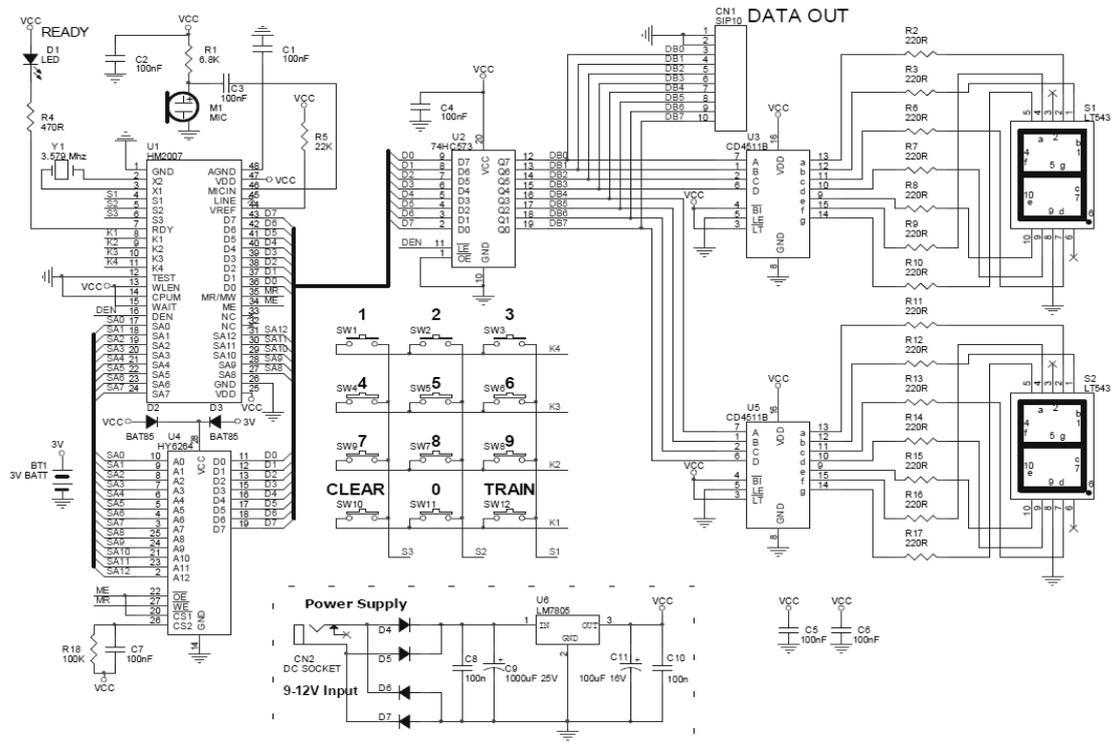


Figure 3 The Complete Schematic of Speech Recognition Circuit

The Complete Schematic of interfacing design is shown below in figure 4

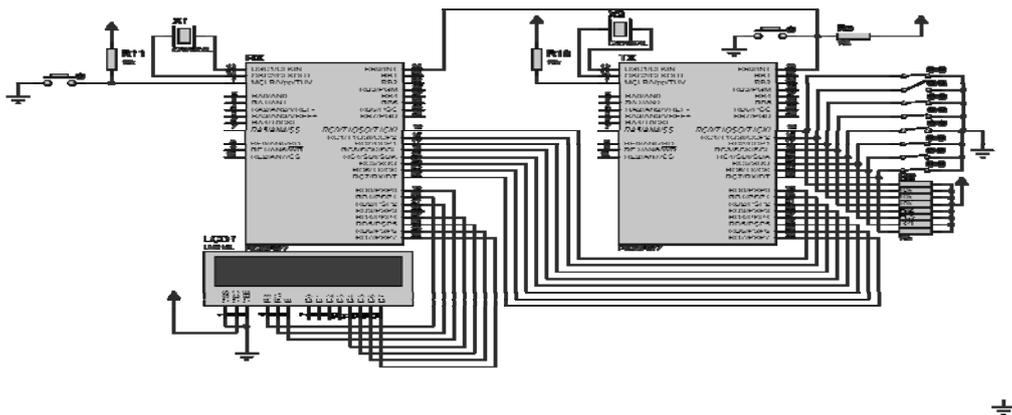


Figure 4 the Complete Schematic of interfacing design

2.2 Training Words for Recognition

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 onboard microphone (near LED) 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 are all you need to train.

2.3 Testing Recognition

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

2.4. Error Codes

The chip provides the following error codes.

55 = word too long

66 = word too short

77 = no match

When interfacing the external circuit through its data bus, The decoding circuit must recognize the word numbers from error codes. So the circuit must be designed to recognize error codes 55, 66 and 77 and not confuse them with word spaces 5, 6 and 7.

2.4 Clearing Memory

To erase all words in memory press “99” and then “CLR”. The numbers will quickly scroll by on the digital display as the memory is erased.

2.5 Changing & Erasing Words

Trained words can easily be changed by overwriting the original word. For instances suppose word six was the word “Capital” and you want to change it to the word “State”. Simply retrain the word space by pressing “6” then the TRAIN key and saying the word “State” into the microphone. If one wishes to erase the word without replacing it with another word press the word number (in this case six) then press the CLR key. In our system we used the following

01	one
02	two
03	three
04	four
05	five
06	six
07	seven
08	eight
09	nine
10	plus

- 11 minus
- 12 multiply
- 13 divide

3. Conclusion

At the end of the design, the system performed most of its functions, although it did not work as a whole. The voice-activated calculator was able to recognize spoken words in natural environments, and was able to run the recognition code on it. The resulting output would get displayed on the LCD in real-time, and the calculations performed by the PIC microcontroller in real-time were also working properly. The time taken for the whole process was very acceptable (about 1 second). However, the major obstacle to successful implementation of the recognition system was the HM2007 itself.

Due to the memory restrictions, the size of the training codebook could not be made larger. Only 10 numbers per classification word were used in this system. But, a larger, more controlled training set could result in higher recognition accuracy.

4. Future Work

For future development of this system, we can implement more SRAM on our speech recognition circuit so that we can recognize more than one digit operation and we can also implement more operation such that log, ln, square root and much more by modifying the code of the PIC controller. We can also implement analog voice output instead of the LCD so it can be used for blind people too. While the voice activated calculator did not work as expected in the demonstration, enough testing has been done to prove that the system is still a feasible one. It is thus imperative to understand from this the importance of making the right engineering decisions for this design.

References

- [1] <http://mi.eng.cam.ac.uk/comp.speech/Section6/Q6.1.html>.
- [2] **Development of Isolated Word Speech Recognition System**, Antennas LIPEIKA, Joana LIPEIKIEN Ę, Laimutis TELKSNYS
- [3] http://en.wikipedia.org/wiki/Speech_recognition
- [4] <http://itp.nyu.edu/physcomp/sensors/Reports/HM2007VoiceRecognitionIC>
- [5] <http://www.pcguides.com/ref/ram/typesSRAM-c.html>
- [6] http://www.datasheet4u.com/html/7/4/L/74LS373_MotorolaInc.pdf.html

- [7] <http://www.doctrionics.co.uk/4511.htm>
- [8] <http://www.datasheetcatalog.org/datasheet/motorola/74LS48.pdf>
- [9] <http://www.embedds.com/matrix-keypad-interfacing-with-microcontrollers/>
- [10] http://www.eidusa.com/Electronics_Voltage_Regulator.htm
- [11] http://en.wikipedia.org/wiki/Crystal_oscillator
- [12] http://www.massmind.org/images/www/hobby_elec/e_pic1.htm
- [13] http://www.futurlec.com/PIC16F877_Controller.shtml

Received: November, 2011