Novateur Publication's
International Journal of Innovation in Engineering, Research and Technology [IJIERT]
ICITDCEME'15 Conference Proceedings
ISSN No - 2394-3696

PSOC BASED ISOLATED SPEECH RECOGNITION SYSTEM FOR VOICE CONTROLLED APPLIANCES

DevarshBharambe
Department of Electronics and Communication
Marathwada Institute of Technology, Aurangabad (MS), India.

Mrs S.P.Kodgire
Associate Professor, Department of Electronics and Communication
Marathwada Institute of Technology, Aurangabad (MS), India.

ABSTRACT
When we say voice control, the first term to be considered is Speech Recognition i.e. making the system to understand human voice. Speech recognition is a technology where the system understands the words (not its meaning) given through speech. The speech recognition system is a completely assembled and easy to use programmable speech recognition circuit. Programmable, in the sense that the words (or vocal utterances) you want the circuit to recognize can be trained. This Programmable System on Chip (PSoC) based embedded system allows you to experiment with many facets of speech recognition technology. Some of interfacing applications that can be made are authentication, controlling industrial motors and hence many other devices connected to it, controlling home appliances, robotics movements, speech assisted technologies and many more.

INTRODUCTION
The human sense of hearing and the human's ability to talk are very important means of communication. Speech recognition will become the method of choice for controlling appliances, toys, tools and computers. At its most basic level, speech controlled appliances and tools allow the user to perform parallel tasks (i.e. hands and eyes are busy elsewhere) while working with the tool or appliance. Speech recognition is the process by which a computer (or other type of machine) identifies spoken words. Basically, it means talking to your computer, and having it correctly recognize what you are saying. This is the key to any speech related application. When we say voice control, the first term to be considered is Speech Recognition i.e. making the system to understand human voice. Speech recognition is a technology where the system understands the words (not its meaning) given through speech. Speech is an ideal method for robotic control and communication. The speech recognition circuit we will outline, functions independently from the machine's main intelligence [Controller]. This is a good thing because it doesn't take any of the controller processing power for word recognition. The controller must merely poll the speech circuit's recognition lines occasionally to check if a command has been issued to the control element. We can even improve upon this by connecting the recognition line to one of the controller's interrupt lines. By doing this, a recognized word would cause an interrupt, letting the controller know a recognized word had been spoken. The advantage of using an interrupt is that polling the circuit's recognition line occasionally would no longer be necessary, further reducing any overhead. Another advantage to this stand-alone speech-recognition circuit (SRC) is its programmability. You can program and train the SRC to recognize the unique words you want recognized. The SRC can be easily interfaced to the controller. To control and command an appliance by speaking to it, will make it easier, while increasing the efficiency and effectiveness of working with that device. 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. Suppose you want to control a menu driven system. What is the most striking property that you can think of? Well the first thought that came to our mind is that the range of inputs in a menu driven system is limited. In fact, by using a menu all we are doing is limiting the input domain space. Now, this is one characteristic which can be very useful in implementing the menu in standalone systems. For example think of the pine menu or a washing machine menu. How many distinct commands do they require? The heart of the circuit is the voice recognition module. ELECHOUSE voice recognition module is a compact and easy-control speaking recognition board. This product is a speaker-dependent voice recognition module. It supports up to 80 voice commands in all. Max 7 voice commands could work at the same time. Any sound could be trained as command. Users need to train the module first before let it recognizing any voice command.

BACKGROUND
The motivation for Isolated Speech Recognition is simple; it is man's principle means of communication and is, therefore, a convenient and desirable mode of communication with machines. Speech communication has evolved to be efficient and robust and it is clear that the route to computer based speech recognition is the

modeling of the human system. Unfortunately from pattern recognition point of view, human recognizes speech through a very complex interaction between many levels of processing; using syntactic and semantic information as well very powerful low level pattern classification and processing.

At the highest level, all speaker recognition systems contain two main modules feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used to represent each speaker. Feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features from his/her voice input with the ones from a set of known speakers. All Recognition systems have to serve two different phases. The first one is referred to the enrollment sessions or training phase while the second one is referred to as the operation sessions or testing phase. In the training phase, each registered speaker has to provide samples of their speech so that the system can build or train a reference model for that speaker. In case of speaker verification systems, in addition, a speaker-specific threshold is also computed from the training samples. During the testing (operational) phase, the input speech is matched with stored reference model and recognition decision is made.

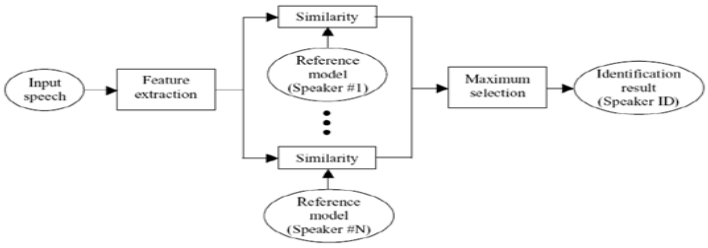

Figure 1: Speaker Identification (Training)

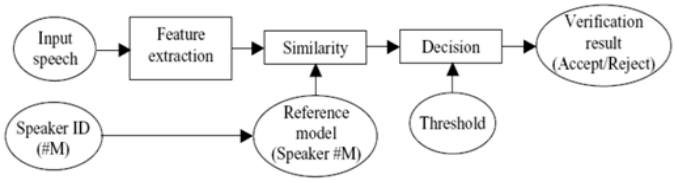

Figure 2: Speaker verification (Testing)

SPEECH RECOGNITION STYLE

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

ISOLATED: This speech recognition system 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 and command spoken. Isolated word recognizers usually require each utterance to have quiet (lack of an audio signal) on both sides of the sample window. It doesn't mean that it accepts single words, but does require a single utterance at a time. Often, these systems have "Listen/Not-Listen" states, where they require the speaker to wait between utterances (usually doing processing during the pauses). Isolated Utterance might be a better name for this class.

CONNECTED: Connect word systems (or more correctly 'connected utterances') are similar to Isolated words, but allow separate utterances to be 'run-together' with a minimal pause between them. It is a half way point between isolated word and continuous speech recognition.

CONTINUOUS: Continuous recognition is the next step. Recognizers with continuous speech capabilities are some of the most difficult to create because they must utilize special methods to determine utterance boundaries. Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. Basically, it's computer dictation. It is the natural conversational speech we are use to in everyday life. It is extremely difficult for a recognizer to shift through the text as the word tends 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.

NATURE OF PROBLEM

Speech recognition is the process of finding a interpretation of a spoken utterance; typically, this means finding the sequence of words that were spoken. This involves pre-processing the acoustic signals to parameterize it in a more usable and useful form. The input signal must be matched against a stored pattern and then makes a decision of accepting or rejecting a match. No two utterances of the same word or sentence are likely to give rise to the same digital signal. This obvious point not only underlies the difficulty in speech recognition but also means that we be able to extract more than just a sequence of words from the signal.

There are a few problems in speech recognition that haven't yet been discovered. However there are a number of problems that have been identified over the past few decades most of which still remain unsolved. The different types of problems we are going to face in our project have been enumerated below:

- **DETERMINING WORD BOUNDARIES:** Speech is usually continuous in nature and word boundaries are not clearly defined. One of the common errors in continuous speech recognition is the missing out of a minuscule gap between words. This happens when the speaker is speaking at a high speed.
- **VARYING ACCENTS:** People from different parts of the world pronounce words differently. This leads to errors in ASR. However this is one problem that is not restricted to ASR but which plagues human listeners too.
- **LARGE VOCABULARIES:** When the number of words in the database is large, similar sounding words tend to cause a high amount of error i.e. there is a good probability that one word is recognized as the other.
- **CHANGING ROOM ACOUSTICS:** Noise is a major factor in ASR. In fact it is in noisy conditions or in changing room acoustic that the limitations of present day ASR engines become prominent.
- **TEMPORAL VARIANCE:** Different speakers speak at different speeds. Present day ASR engines just cannot adapt to that.

SYSTEM MODELLING

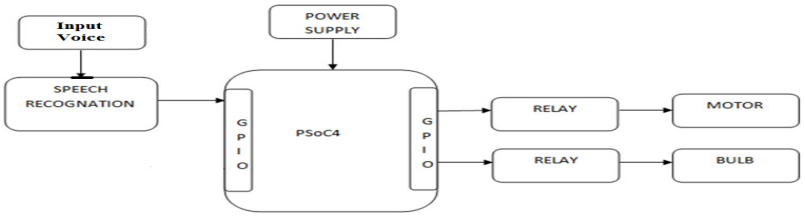

Figure 3: Basic Block Diagram

The system consists of PSoc4, Relays, LCD, motor, bulb, Speech recognition kit and a mic. Speech recognition kit is used to get a speech input and compare the input speech with database.

A. VOICE RECOGNITION MODULE

It is the heart of the entire system. Voice recognition chip with on-chip analog front end, voice analysis, and recognition process and system control functions. The input voice command is analyzed, processed, recognized and then obtained at one of its output port which is then decoded , amplified and given to PSoC.

B. PROGRAMMABLE SYSTEM ON CHIP (PSOC)

PSoC (Programmable System on Chip) is a family of microcontroller integrated circuits by Cypress Semiconductor. These chips include a CPU core and mixed signal arrays of configurable integrated analog and digital peripherals.

PSOC DEVICES FEATURES:

- Configurable Analog Blocks
  - Implement ADCs, DACs, filters, amplifiers, comparators, etc.
- Configurable Digital Blocks
  - Implement timers, counters, PWMs, UART, SPI, IrDA, etc.
- 4KB to 32KB of Flash memory for program storage

- 256B to 2KB of SRAM for data storage
- M8C Microcontroller: 4 Million Instructions Per Sec.

CONCLUSION

From above explanation we conclude that voice recognition module can be used to detect voice signals accurately. After detecting voice signals can be used to operate the motors and light system. Thus, we can implement Programmable System on Chip (PSoC) isolated speech recognition system for human machine interface in embedded system.

REFERENCES

[1] V. Naresh, B. Venkataramani, Abhishek Karan and J. Manikandan.,"PSoC based isolated speech recognition system" International conference on Communication and Signal Processing, April 3-5, 2013
[2] SantoshK.Gaikwad,BhartiW.Gawali,PravinYannawar," A Review on Speech Recognition Technique" International Journal of Computer Applications Volume 10– No.3, November 2010
[3] M.A.Anusuya,S.K.Katti "Speech Recognition by Machine: A Review",International Journal of Computer Science and Information Security,Vol. 6, No. 3, 2009
[4] Dr. SunitaRana "The process of speech recognition, perception, speech signals and speech production in human beings", International Journal of Advanced Research in Computer Engineering & Technology Volume 1, Issue 9, November 2012
[5] Nidhi Desai, Prof.KinnalDhameliya, Prof.Vijayendra Desai "Feature Extraction and Classification Techniques for Speech Recognition: A Review", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 12, December 2013
[6] Ms. Vrinda ,Mr. ChanderShekhar "Speech recognition system for English language",International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013
[7] Sonali N. Malshikare, Prof. V. M. Sardar "Speech Enhancement through Elimination of Impulsive Disturbance Using Log MMSE Filtering", International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015
[8] LindasalwaMuda, MumtajBegam and I. Elamvazuthi "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010
[9] Shipra J. Arora ,Rishi Pal Singh "Automatic Speech Recognition: A Review", International Journal of Computer Applications (0975 – 8887) Volume 60– No.9, December 2012
[10] SadaokiFurui "50 Years of Progress in Speech and Speaker Recognition Research" ECTI transactions on computer and information technology vol.1, no.2 november 2005
[11] James M. Kates, Senior Member, IEEEA Time-Domain Digital Cochlear Model", IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL 39. NO. 12, DECEMBER 1991
[12] By L. R. RABIN ER and M. R. SAMBUR"An Algorithm for Determining the Endpoints of Isolated Utterances"
[13] Sven E. Kr¨uger, Martin Schaff ¨oner, Marcel Katz, EdinAndelic, Andreas Wendemuth"Speech Recognition with Support Vector Machines in a Hybrid System" IESK, Cognitive Systems, Otto-von-Guericke-University Magdeburg,PF 4120, 39016 Magdeburg, Germany
[14] Nobuo Hataoka ,YasunariObuchi, Teruko Mitamura, Eric Nyberg"Robust Speech Dialog Interface for Car Telematics Service"
[15] AshutoshSaxena and Abhishek Singh "A Microprocessor based Speech Recognizer for Isolated Hindi Digits"
[16] http://www.cypress.com/products/programmable-system-chip-psoc
[17] http://artsites.ucsc.edu/EMS/music/tech_background/TE-20/teces_20.html#I.
[18] www.Howstuffworks.com for understanding microphone concepts and other related concepts.
[19] Lawrence Rabiner, Biing-Hwang Juang "Fundamentals of Speech Recogn- ition"