

Implementation of an arm9 board to be controlled by speech using Qtopia on Linux

Chandni K Kella¹, Prof. Srushti P Soni²

¹(EC Department, Gujarat Technological, University, India)

²(EC Department, Gujarat Technological University, India)

Abstract : Voice recognition applications are becoming more useful nowadays. With growth in the needs for embedded computing and the demand for emerging embedded platforms, it is required that voice recognition systems are available but voice recognition software being closed source cannot be used easily for implementation of voice recognition based devices. To implement the approach on a real-time application, an ARM 9 device interface is to be designed to control the applications on Linux platform, it being an open source using Qtopia software.

Keywords: Qtopia, voice recognition, lexical decoding, phenomes

I. INTRODUCTION

Human-Embedded device voice interface has a key role in many application fields. It is proposed that a new approach based on the recognition of isolated words, using a set of traditional pattern recognition approaches and a discrimination approach based on test results of classical methods and in order to increase the rate of recognition. The increase in complexity as compared to the use of only traditional approach is negligible, but the system achieves considerable improvement in the matching phase, thus facilitating the final decision and reducing the number of errors in decision taken by the voice command guided system. Moreover, speech recognition constitutes the focus of a large research effort in Artificial Intelligence (AI), which has led to a large number of new theories and new techniques.

The application is based on the voice command; it therefore involves the recognition of isolated words from a limited vocabulary used to control the device.

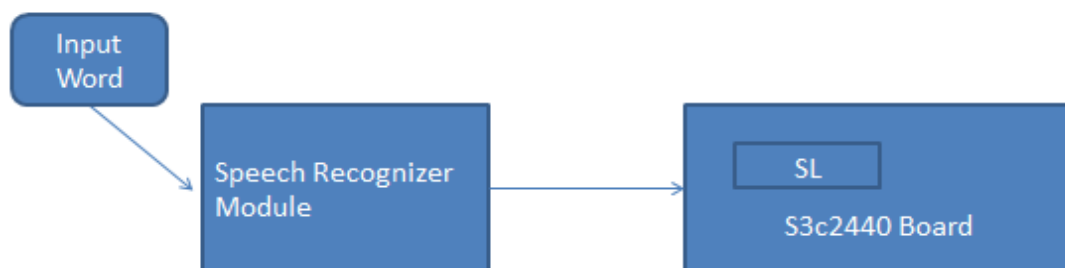


Fig 1.1 Basic block diagram

The speech recognition agent is based on a traditional pattern recognition approach. The main elements are shown in the block diagram of Figure 1.1. The pre-processing block is used to adapt the characteristics of the input signal to the recognition system. It is essentially a set of filters, whose task is to enhance the characteristics of the speech signal and minimize the effects of the background noise produced by the external conditions.

The Speech locator (SL) block detects the beginning and end of the word pronounced by the user, thus eliminating silence. It processes the samples of the filtered input waveform, comprising useful information (the word pronounced) and any noise surrounding the PC. Its output is a vector of samples of the word (i.e.: those included between the endpoints detected).

II. SPEECH RECOGNITION SYSTEM

Speech recognition basically means talking to a computer, having it recognize what we are saying, and lastly, doing this in real time. This process fundamentally functions as a pipeline that converts PCM (Pulse Code Modulation) digital audio from a sound card into recognized speech.

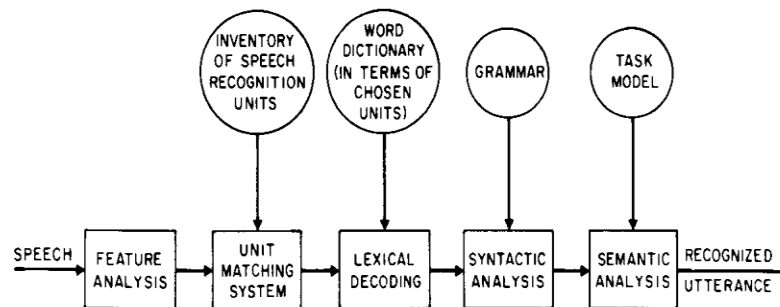


Fig1.2 Block diagram of speech analysis

Transform the PCM digital audio into a better acoustic representation – The input to speech recognizer is in the form of a stream of amplitudes, sampled at about 16,000 times per second. But audio in this form is not useful for the recognizer. Hence, Fast-Fourier transformations are used to produce graphs of frequency components describing the sound heard for 1/100th of a second. Any sound is then identified by matching it to its closest entry in the database of such graphs, producing a number, called the “feature number” that describes the sound.

Unit matching system provides likelihoods of a match of all sequences of speech recognition units to the input speech. These units may be phones, diphones, syllables or derivative units such as fenones and acoustic units. They may also be whole word units or units corresponding to group of 2 or more words. Each such unit is characterized by some HMM whose parameters are estimated through a training set of speech data.

Lexical Decoding :constraints the unit matching system to follow only those search paths sequences whose speech units are present in a word dictionary.

Apply a "grammar" so the speech recognizer knows what phonemes to expect. This further places constraints on the search sequence of unit matching system. A grammar could be anything from a context-free grammar to full-blown English.

Figure out which phonemes are spoken – This is a quite dicey task as different words sound differently as spoken by different persons. Also, background noises from microphone make the recognizer hear a different vector. Thus a probability analysis is done during recognition. A hypothesis is formed based on this analysis. A speech recognizer works by hypothesizing a number of different "states" at once. Each state contains a phoneme with a history of previous phonemes. The hypothesized state with the highest score is used as the final recognition result.

A. RELATED WORK

A lot of speech aware applications are already there in the market. Various dictation softwares have been developed by Dragon, IBM and Philips. Genie is interactive speech recognition software developed by Microsoft. Various voice navigation applications, one developed by AT&T, allow users to control their computer by voice, like browsing the Internet by voice. Many more applications of this kind are appearing every day .The SPHINX speech recognizer of CMU provides the acoustic as well as the language models used for recognition. It is based on the Hidden Markov Models (HMM). The SONIC recognizer is also one of them, developed by the University of Colorado. There are other recognizers such as XVoice for Linux that take input from IBM’s Via Voice which, now, exists just for Windows. Background noise is the worst part of a speech recognition process. It confuses the recognizer and makes it unable to hear what it is supposed to. One such recognizer has been devised for robots that, despite of the inevitable motor noises, makes it communicate with the people efficiently. This is made possible by using a noise-type-dependent acoustic mode l corresponding to a performing motion of robot. Optimizations for speech recognition on a HP Smart Badge IV embedded system has been proposed to reduce the energy consumption while still maintaining the quality of the application. Another such scalable system has been proposed in for DSR (Distributed Speech recognition) by combining it with scalable compression and hence reducing the computational load as well as the bandwidth requirement on the server. Various capabilities of current speech recognizers in the field of telecommunications are described in like Voice Banking and Directory Assistance.

Voice Input: The input is human voice which, as explained before, is sampled at rate of 16,000 per second. It should be given in live mode. But because of some conflicts in the channel settings of the sound card and that used by the software, we are not able to do it in live mode. We are running the recognizer in batch mode, instead, i.e. taking input in the form of a pre-recorded audio file (in RAW format).

Microphone: The microphone that we are using for recognition is built onto the PXA27xplatform itself. It has got its own advantages and disadvantages:

Advantages:

- Nothing to plug in.
- User's hands are free.

Disadvantages:

- Low accuracy unless the user is close to the monitor.
- Not good in a noisy environment.

B. TYPES OF VOICE RECOGNITION SYSTEM

- **Speaker dependent system** - The voice recognition must be trained before it can be used. This often requires a user reads a series of words and phrases so the computer can understand the users voice.
- **Speaker independent system** - The voice recognition software recognizes most users voices with no training.
- **Discrete speech recognition** - The user must pause between each word so that the speech recognition can identify each separate word.
- **Continuous speech recognition** - The voice recognition can understand a normal rate of speaking.
- **Natural language** - The speech recognition not only can understand the voice but also return answers to questions or other queries that are being asked.

III. Simulation Results

The snapshots below describe the results of text to speech conversion using Qt creator using Flite algorithm.

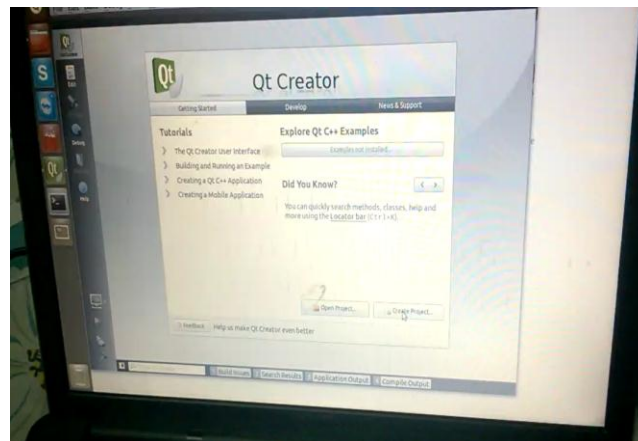


Fig.1.3 QT creator window

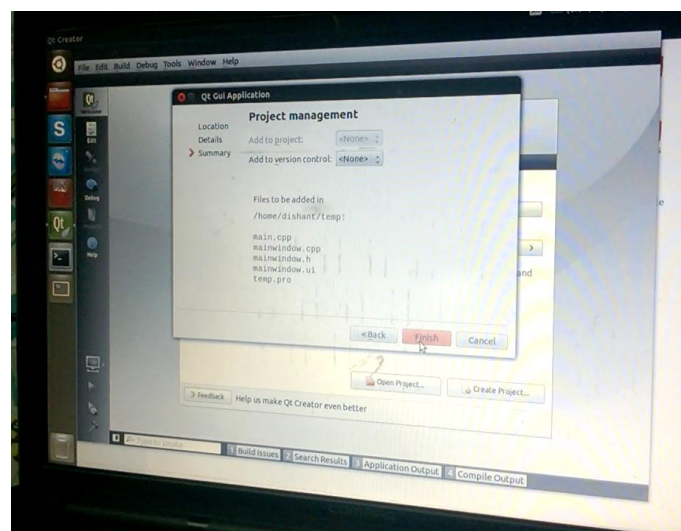


Fig .1.4 Click on finish and a new project will be created

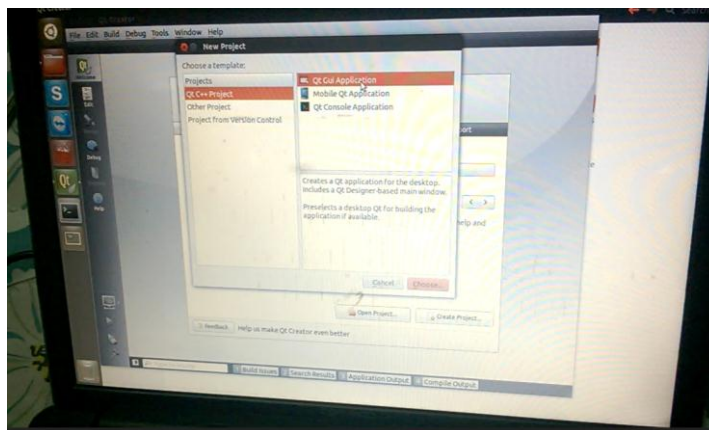


Fig. 1.5 QT c++ project is to be selected, in which we will have to go for QT Gui application.

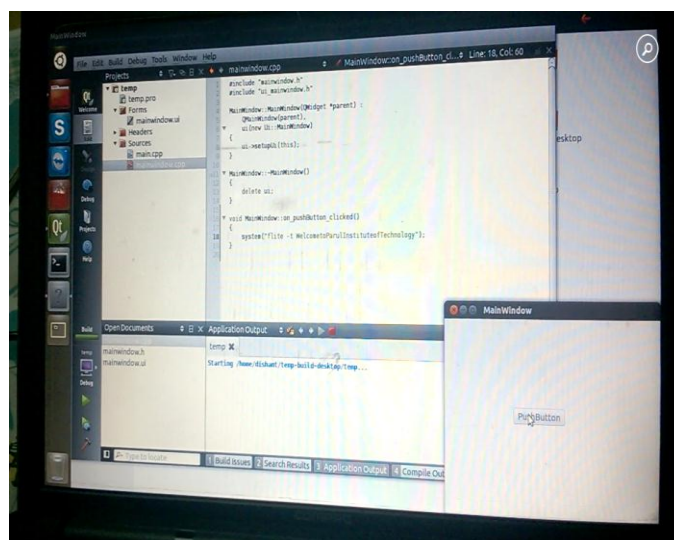


Fig. 1.6 we will get the output where or text will be converted to speech. We will click on push button and get the desired output in form of speech.

IV. CONCLUSION

From my research I conclude that I will be using open source QTOPIA software based on LINUX platform to control an arm9 device through voice commands. I have developed two applications which are controlled through voice command.

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to all of them. I am highly indebted to **Prof. Srushti P. Soni, PIET, Limda** for her guidance and constant supervision as well as for providing necessary information regarding the project. I would like to thank IOSR journal for the support to develop this unique document.

REFERENCES

- [1] Implementation of a voice communication based ARM9 device Author : Istvan I. Papp, *Member*, IEEE, Zoran M. Šarić, Nikola Dj. Teslić, *Member*, IEEE Year of publication: IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, May 2011
- [2] A HYBRID METHOD FOR AUTOMATIC SPEECH RECOGNITION PERFORMANCE IMPROVEMENT IN REAL WORLD NOISY ENVIRONMENT Author : **Urmila Shrawankar and Vilas Thakare** Department of Computer Science and Engineering, G H Raisoni College of Engineering, Nagpur, India Department of Computer Science and Engineering, SGB Amravati University, Amravati, India Year of publication: **Journal of Computer Science, 2013 publication**
- [3] Speech in Mobile and Pervasive Environments - By Nitendra Rajput, Amit Anil Nanavati (eBook)
- [4] Sphinx-4: A Flexible Open Source Framework for Speech Recognition, Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, Joe Woelfel
- [5] H. Chung and I. Chung, "Memory efficient and fast speech recognition system for low-resource mobile devices," IEEE Trans. Consum. Electron., vol. 52, no. 3, pp. 792-796, 2006.
- [6] M. Ji, S. Kim, H. Kim, and H.-S Yoon, "Text-independent speaker identification using soft channel selection in home robot environment," IEEE Trans. Consum. Electron., vol. 54, no. 1, pp. 140-144, 2008.