ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 1, November 2012

Text to Speech Conversion Using Raspberry-Pi for Embedded System

P.V.N. Reddy

Professor, Department of ECE, S V College of Engineering, Tirupati, A.P, India

ABSTRACT: This paper proposes a system which is used for converting the input string of text into the corresponding speech using Raspberry-pi. The fastest and effective way of communication is language. Limited and proper combination of words with grammar rules gives a clear picture of the ideas or thoughts that speaker wants to convey. The system includes Python coding which is done on Raspberry-pi for the generation of speech signal based on the user defined input text. A simple File Accessing Protocol (FAP) is adapted to achieve the task of retrieving the audio files on the database. This paper also throws light on the mechanism and strategy used to convert the text input to speech output.

KEYWORDS: Raspberry-pi, Text to speech, Speech generation, Speech synthesis.

I. INTRODUCTION

Text to speech is a system which is based on the automatic generation of the new sentences or words. This is a method where a computer is made to speak. The dictionaries are often referred for the correct pronunciation of words but for some of the words the dictionaries cannot give the correct pronunciation. Thus the dictionary is only limited for a set of words. The proposed system is just limited to some of the audio conversion that user wants to listen since the dictionary is also limited to some set of words thus it is not possible to cover entire word string.

As the paper discusses about text to speech conversion speech synthesis comes into frame. Speech synthesis is a artificial or computer generated human speech. A system which is used for this purpose is called speech synthesizer. A rule based text-to-speech synthesis system was developed for the language of Malaysia known as Malay [1]. This speech synthesis system also known as SMaTTS, was developed with user-friendly graphical user interface (GUI).The algorithms for SMaTTS system were developed and were compared. The overall performance of the system is analysed using categorical estimation (CE) for a comprehensive analysis.

Another speech synthesizer was developed for the Macedonian language [2] which is based on concatenation of speech segments. Concatenation techniques has overcome the barrier of obtaining the extra-linguistic information by using the actual voice of a recognizable person as a source, combined with minimal use of signal processing. But still there is a problem when the person speaks in various prosodies. Speech synthesis can also be done by utilizing glottal inverse filtering for generating natural sounding speech [3].

Speech synthesisation can be performed in number of ways one of them is Unit Selection Synthesis which provides with more naturalness and efficiency based on that unit is used like phoneme, diphone, syllable etc[5]. A text-to-speech system was developed for converting Thai text to speech (TTTS). TTTS consists of four modules including thai text, text processing, dictionary lookup and speech synthesis. The TTTS dictionary contains around 5300 thai words along with their sound wave[6]. Neural networks can also be used for text to speech conversion. Use of neural networks can absorb many of the irregularities in Greek pronunciation[8]. Chinese
language is analysed based on the analysis of characteristics of Tibetan language[9]. SAMPA_ST standard pronunciation is being used for this purpose. A text-to-speech system for Azerbaijani Turkish language was developed which explains various speech synthesis methods and some important features of Azerbaijani Turkish language. Corpus based concatenative speech synthesis is used for this system. According to the input text appropriate units are selected and then concatenated[10].

Automatic speech recognition (ASR) transforms speech into text [4]. ASR is developed based on speech coding theory, while simulating certain spectral analyses performed by the ear. Speech synthesis is the automatic generation of a speech waveform, typically from an input text. TTS makes use of database of information, both speech and text. Previously analysed speech is stored in small units in the database, for concatenation in the proper sequence at runtime. TTS systems
first perform text processing, including "letter-to-sound" conversion, to generate the phonetic transcription. The database containing words or sentences gives higher quality of outputs. Furthermore, a synthesizer can also incorporate human voice characteristics to create a synthetic
but somewhat natural voice. Mainly a speech synthesizer is judged on the basis of the similarity of output speech obtained with the human voice [5]. Upcoming days Text to speech (TTS) system focuses more on emotional

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 1, November 2012

proportion with the speech i.e. it requires more realistic and natural voice that can be closely matched with the human voice. Thus there comes a limitation in input text in order to get high quality of output. The examples for speech synthesis are voice enabled e-mail and unified messaging. The first step involved in speech synthesis is for the user to speak a word in microphone than the speech is digitized using Analog to Digital (A/D) converter and it is stored in memory in the form of database.

In order to identify this voice the computer matches it with the voice sample that has some known meaning. Here comes another application of controlling any device with voice rather than using keyboard or mouse . Text to speech conversion basically involves two parts front end and back end. The front end process involves language related processing like Text tokenization, Text to speech conversion etc [2]. The back end process involves the speech related signal processing like pitch estimation, speech synthesis etc. Normally the front end steps remain the same while back end steps varies because of the different synthesizers involved based on the output desired by the user. Text tokenization refers to the words of the string separated by space, commas etc. Text normalization basically refers to the conversion of numbers, abbreviations into

corresponding words, for e.g. Mr., Ms., Dr.etc. Apart from this front end steps also involves parts of speech tagging, phrasing, pause insertion etc. the back end step involves the tone modulation and the generation of speech based on some algorithm. The whole process of Text to speech conversion is carried out in Raspberry-pi which contains an ARM processor which further gives the high performance and accurate results.

II. RASPBERRY-PI

The Raspberry-pi is a small size computer which does not have hard disk as there in PC but has an SD card of 4GB or above for storage purpose. There are numerous interactive application of this small size computer from high quality audio, video playback to playing 3D games etc [4]. It has inbuilt ARM processor in it. ARM processors can be considered as the brain of Raspberry-pi. They are highly efficient in solving complex algorithm in less fraction of time. That is why they are used in small devices like mobile phones, gaming devices and other digital devices. The OS for Raspberry should be installed in SD card apart from this it should store all the program files needed by Raspberry. The SD card should be formatted before installing the OS. The OS

should be burnt using Windows 32 disk imager. The OS in the Raspberry acts as the interface between user and the Raspberry [4]. Fig.1 shows Model B+ of Raspberry-pi.

Fig. 1 Architecture of Raspberry-Pi

As shown the Fig. 1 Model B+ of Raspberry consists of four ports to connect hardwares like mouse, keyboard, web cam etc. It has one HDMI port which is used to connect from monitor or TV and one Ethernet port for connectivity from internet. It has one micro USB power port to power on the Raspberry and one audio port. RCA video port is provided for the connectivity from older type of TV. Apart from this it consists of GPIO headers to connect from other hardware devices like LED, motor etc. The proposed project makes use of GPIO ports. Since the

keyboard is connected to Raspberry-pi, its GPIO pins needs to be initialize. Raspberry-pi is advantageous for text-to-speech conversion in many ways like: It provides various alternatives for speech synthesis like eSpeak, Festivals, fite etc. Raspberry-pi draws a very less amount of power around 5-7 Watts. It has expansion capabilities from GPIO to

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 1, November 2012

camera etc. The HDMI display port can handle the resolution up to 1920X1200 which can be further used for some future applications.

III. RESULTS & DISCUSSION

The existing project consists of Raspberry-Pi, PC and a speaker. Python language is used for coding the Raspberry-pi. There are number of ways though which the input text can be converted into speech like string matching, frequency matching etc. Initially, the GPIO pins should be initialised in order to use them as input or output. This can be coded in python by importing GPIO module. At first the coding is done for matching the English alphabets with the user defined alphabets and playing the corresponding audio file. Audio files can be database or voice recordings. Python coding provides one more logic called frequency matching. In this method the frequencies on the basis of repetition of alphabets for each word should be written and then as per the coding logic all the words having matched frequencies should be called and corresponding audio for that word should be played.

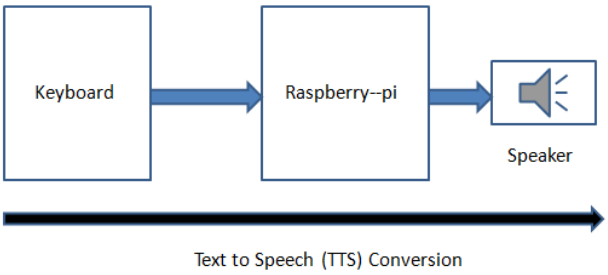

Fig. 2 Block diagram implementation of TTS system.

Fig. 2 shows the block diagram for Text to Speech (TTS) conversion. Another method for converting the text into speech can be through the ASCII values of English letters. By using this method the coding length can be decreased. There are many Text to Speech converters are there but there performance depends on the fact that the output voice is how much close to the human natural voice. For example, consider a name pretty, it can be a name of a person as well as it can be defined as beautiful. Thus it depends on how the words are pronounced. Many text to speech engines does not give the proper pronunciation for such words thus combining some voice recordings can give more accurate result. The TTS system converts an English text into a speech signal with prosodic attributes that improve its naturalness. There are many systems which include prosodic processing and generation of synthesized control Parameters. The proposed system provides good quality of synthesized speech. The text processing component provides reliable grammatical classification and phonetic transcription. Fig. 3 shows the flow chart diagram for TTS system. Fig. 4 shows a pseudo code written in python for converting English letters into speech.

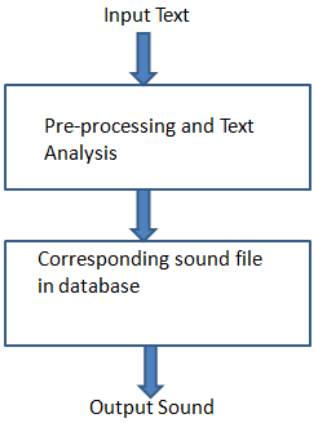

Fig. 3 Flow Chart Diagram

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 1, Issue 1, November 2012

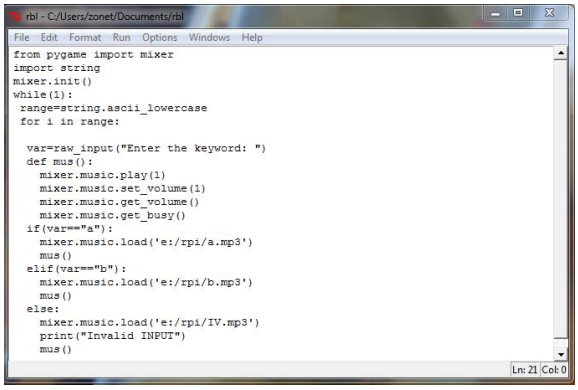

Fig 4. Pseudo code of the program implemented using Python

Prototype implementation of the proposed system is done. Fig. 5 shows the image of the Text to speech (TTS) conversion. The input string or word is given by the user to Raspberry-pi which further uses python coding and some programming logic, calls the corresponding audio file of that input text. Thus, Text to Speech conversion is done successfully for the set of words by using some algorithm. Fig. 6 shows the hardware setup which includes Raspberry-pi connected to the computer. The work can be extended for using the video characteristics of Raspberry-pi to make the proposed system more effective. The further application can also be extended to controlling the electronic devices like LED through speech signal.

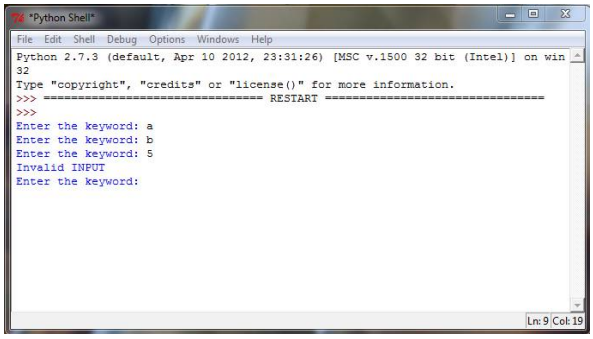

Fig. 5 Console output for Text to Speech Conversion

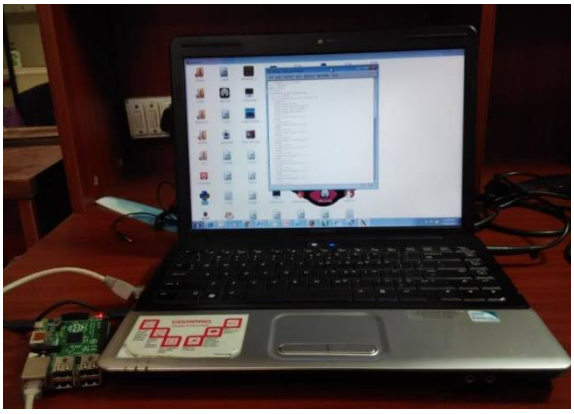

Fig 6. Interface of Raspberry-Pi TTS System connected to PC

IV. CONCLUSIONS

The proposed Text to speech (TTS) system was designed in order to produce an equivalent acoustic signal which goes in synchronization with the text which is provided as an input to the Raspberry-Pi system. The programming strategy involved the GPIO lines, the database connectivity and the audio amplifiers. This experimental validation can be extended to conjoined sounds and words of higher complexity also. The scope of the Raspberry-Pi in speech synthesis and feature extraction is also quite boundless.

REFERENCES

[1] Ahmad, Z.H. ; Int. Islamic Univ. Malaysia, Kuala Lumpur ; Khalifa, O., "Towards designing a high intelligibility rule based standard malay text-to-speech synthesis system" IEEE, International Conference on Computer and Communication Engineering. pp 89-94, May 2008.
[2] Chungurski, S. ; Fac. of Commun. & Inf. Technol., FON Univ., Skopje ; Kraljevski, I. ; Mihajlov, D. ; Arsenovski, S. "Concatenative speech synthesizers and speech corpus for Macedonian language" International Conference on Information Technology Interfaces. ITI 2008., pp 669-674, June 2008.
[3] Raitio, T. ; Dept. of Signal Process. & Acoust., Aalto Univ., Helsinki, Finland ; Suni, A. ; Yamagishi, J. ; Pulakka, H. , "HMM-Based Speech Synthesis Utilizing Glottal Inverse Filtering" IEEE Transactions on Audio, Speech, and Language Processing, pp 153-165, Oct 2010.
[4] *www.raspberrypi.org*.
[5] Yagi, Y. ; Graduate Sch. of Eng., Tokyo Univ. ; Hirose, K. ; Takada, S. ; Minematsu, N., "Improved concept-to-speech generation in a dialogue system on road guidance" Cyberworlds ,IEEE, International Conference , pp 8 - 436, May 2005.
[6] Sadeque, F.Y.; Dept. Of Comput. Sci. & Eng., Yasar, S. ; Islam, M.M., "Bangla text to speech conversion: A syllabic unit selection approach" Cyberworlds ,IEEE, International Conference , pp 1-6, May 2013.
[7] Pornpanomchai, C., Soontharanont, N. ; Langla , C. ;Fac. Of inf. & Commun. Technol. ,Mahidol Univ., Bangkok;, "A Dictionary-Based Approach for Thai Text to Speech (TTTS)"Third International Conference on Measuring Technology & Mechatronics Automation ,IEEE, International Conference , pp 40-43, Jan 2011.
[8] Falas.T., Stafylopatis, A-G ;Dept. Of comput. Sci. & Eng. , Cyprus coll., Lefkosia, Cyprus; ; Graduate Sch. of Eng., Tokyo Univ. ; Hirose, K. ; Takada, S. ; Minematsu, N., "Neural networks in text-to-speech systems for the Greek language", Electrotechnical Conference, 2000. MELECON 2000. 10 th Mediterranean, pp 574-577, 2000.
[9] Lu Gao, Hongzhi Yu, Yonghong Li, Jie Liu.; Key Lab of China's Nat. Linguistic Inf. Technol., Northwest univ. for Nat., Lanzhou, China. "A Research on Text analysis in Tibetan Speech sysnthesis"Information And Automation(ICIA),2010, IEEE International Conference , pp 817 - 822, June 2010.
[10] Damadi, M.S., Azami, Bahram Zahir ; Electr. & Comput. Eng. Dept., Univ. of Kurdistan, Sanandaj, Iran, "Design and Evaluation of a Textto-Speech System for Azerbaijani Turkish Language and Database Generation" Digital Telecommunications,2009.ICDT'09. , pp 111 - 116, July 2009.