International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

Embedded Based Implementation of Real Time Text-To-Speech Conversion

Shantinath. A. Sanaki[1]; Bharath B.S.R[2] & Mr. V. Nattarasu[3]
Student, Electronics and Communication Engineering, SJCE, Mysuru, India[1]
Student, Electronics and Communication Engineering, SJCE, Mysuru, India[2]
Assoc. Professor, Electronics and Communication Engineering, SJCE, Mysuru, India[3]

Abstract:
"Speech" and "gestures" are the two expressions, which are used in communication between human beings. In human communication, the use of speech and gestures is coordinated. Real time hardware implementation of Text to Voice system has been drawing attention of the research community due to its various real time applications. The main objective is to provide any valid text at the input via keyboard and the output will be obtained as a voice representation of the input text, which includes displaying of input text as we type and wireless transmission of typed data over a distance. This will be helpful for the visually impaired persons, taking aid for the vocally disabled, training aids and other commercial applications. All these applications demand the real time embedded platform to meet the real time specifications such as speed, power, space requirements etc.
Previous approach used was gesture to speech conversation. That is a sign language recognition, which deals with recognition of hand gesture. This was implemented by wearing hand glows, to get the figure variations. The technique used for gathering information is: Image based or device based, that is by using camera, detection of light variations or by using instrument gloves i.e. position tracker. But in a continuous stream it is challenging task to recognize the sign.
In this context, proposed approach is the use of embedded processor to implement Text-To-Speech conversion. LPC2148 microprocessor has been chosen as the hardware platform which is an Advanced RISC Machine, message is displayed on LCD and transmitted using radio frequency modulator. Two modules are built for the prospective of transmitter and receiver. The hardware is rooted with C code, which is compiled in Keil and programmed to controller.

Keywords: ARM; Text-To-Speech; keyboard; visually impaired

I. Introduction

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer. This can be implemented either in software or hardware domain. A text-to-voice system converts normal language text into speech. It is microcontroller based hardware, coded in Embedded C language.

Well articulated text-to-speech program, allows people with visual disabilities to listen. Currently, dumb people use hand language to convey their message to others, but listener is not able to capture all information, this creates barrier between people. Present system's aim is to make an easy conversion from text to speech.

The operands that are to be compelling recognition are the alphabets and the numbers. The main intended goal to be attained is to provide any valid text at the input via keyboard and the output is produced as a voice representation of the input text.

This is based on the transition of text into speech using the main frame speech processor. When the desired text is provided as an input to the controller, it transmits the text to the controller via PIC which sends stored scan codes to controller which processes the data, stores in its memory i.e. into buffer as the ASCII codes. The data is transferred on air using Xbee transmitter and at the receiver side same data is captured by Xbee receiver which in turn will be directly connected to the controller. Received data is temporarily stored in the buffer, according to clock rate scan codes are transmitted to TTS controller which is finally connected to SpeakJet which produces speech via speaker.

A. *Motivation*

In this world there are people those who are not able to see, speak and hear properly. According to census statistics the disability rate is increasing year by year [7]. These people will be facing difficulty in conveying information with normal persons. If they meet the person who don't know the sign language then it is very difficult to communicate.

This prompted to come up with a new idea that is, the person who cannot talk, can type a message that they want to share and the listener can hear the speech output or can see the message on display. If the person is visually impaired, they need to get trained with typing text on keyboard and can analyze the typed text as they hear the speech.

The scope of this project is big, in terms of cost, facility and our contribution lies in selecting the suitable microcontroller architecture, interfacing various ICs with the microcontroller after studying their datasheets. The advantage of our proposed project is to help visually impaired people in their daily life. The proposed method is the better system compared to gesture recognition and easy to implement.

B. *Problem Statement*

The people who are visual impaired and the people who are not able to speak and hear properly, will use sign language to communicate. For this, the project for sign to speech conversion, by using hand gesture movements can be developed. But we get very few outputs, since the movements we can make by hand, are less and we cannot share all the words with limited signs and the communication barrier will occur with people who do not understand sign languages.

Instead of sending few responses by hand gesture, an electronic controller can be used. The use of electronics controller is for effective communication. This can be achieved through text to speech conversion with display, there by effecting complete convey of message.

C. *Approach To Solution*

Scan codes are found for the alphabets and numerical values. Determined scan codes are stored in the PIC and compared with input data using case statement. To interface keyboard to processor, PIC is used as a bridge. By defining the port, the data is transmitted to RF module and to the processor. For all transfer and receive functions, indication should be given accordingly. Display alignment is controlled by processor and displayed as the data received. The receiver system is built with the same microprocessor, display and includes text to speech controller. The TTS should be interfaced and controlled by the processor. At the end speaker interfaces to the TTS controller module.

II. LITERATURE REVIEW

For the last 35 years research on TTS has been carried out extensively, to explore the essential features. This section presents an extensive literature review of past and present research on the different distinct features of the TTS system [1].

Generating synthetic speech has been somewhat unusual in the past. Looking at the recent research trends, data glove is discussed in the system design for hand gesture recognition [5], where the system considers the only the single handed gestures.

The disabled person wears the hand glow, the system takes input from this hand glove. The gesture recognition engine uses this input to compare the captured data and the database. The hand glows position and orientation are obtained by an arm cover sensor and data glove. The glow consists of 5 potentiometers each connected to a finger. The variation in resistance gives the particular output as defined in the database.

Each gesture has its own output to the system. Each symbol is numbered from a to z and 0 to 9. The hand movements result in change in potentiometer resistance, these resistance values are stored in the database as a separate address line and each address line is assigned with a vocal message. Each finger movement gives a different message to listener. It can be displayed on a screen or speech can be generated.

The first known mechanical talking machine was introduced by Gerbert of Aurillac [6]. Then inventors like Roger Bacon and Albertus Magnus engaged in creating machine called “talking heads”. However, the first known machine that tried to constituting an imitation of real human speech was formulated by Christian Kratzenstein of St. Petersburg. The machine could produce 5 long vowel sounds. After a few years later, Wolfgang Von Kempelen developed a machine [8] that could produce vowels and some consonants.

At the initial stage of 20th century with increasing use of electricity, speech synthesis began to move from mechanical machines to electrical machines. Then an Electronic speech synthesizer was developed. In the late 1930’s by Homer Dudley developed the 1st known electronic synthesizer called VODER. He was a research physicist at the Bell Laboratories in New Jersey. Dudley reconstructed the Bell Laboratories with speech investigation of the component parts of a whole and their relations in making up the whole equipment called VOCODER. The Voder was a simplified version of the Vocoder i.e. voice encoder.

The VODER stands for “Voice Operation DEmonstratoR”. It was controlled by an operator using a keyboard to adjust the filter output, foot pedals to control the fundamental frequency, and special keys to create closure and release required for stops. The VODER swas operated like a musical instrument. Eventually, human operated machines became obsolete. After World War II, the spectrograph provided a new tool for researching acoustic phonetics. As a result, researchers began to study speech based on acoustic data.

In the 1950’s, speech synthesizers like the Pattern Playback were developed to produce speech from copying speech waveforms. The speech synthesis by rule approach began to become more frequent in the following decade [4]. Concatenate speech synthesis became a focus of research, with the initial focus on phoneme concatenation. In 1976, Olive and Spickenagle used linear prediction speech analysis to automatically create a full diaphone inventory for concatenation. After 10 years, Nakajima and Hamada wrote about a method of speech concatenation that used a unit-selection based approach, instead of the more common diaphones concatenation approach. Today, diaphone concatenation and unit-selection concatenation are the most common methods of speech synthesis with the latter becoming more common commercially.

III. HARDWARE REQUIREMENTS

This section explains in detail the hardware and software required to build the system.

A. *Arm7 (Lpc2148) Microcontroller*

ARM micro-controller is one of the widely used controllers in embedded domain. The ARM architecture is similar to Reduced Instruction Set Computer (RISC) architecture, as it incorporates these typical RISC architecture features:

Figure 1 highlights the ARM7 controller. It is a 32-bit or a 16-bit TDMI-S processor, which supports 16 bit Thumb instruction. A uniform register file load/store architecture, where data processing operates only on register contents, not directly on memory contents.

Simple addressing modes, with all load/store addresses determined from register contents and instruction fields only.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

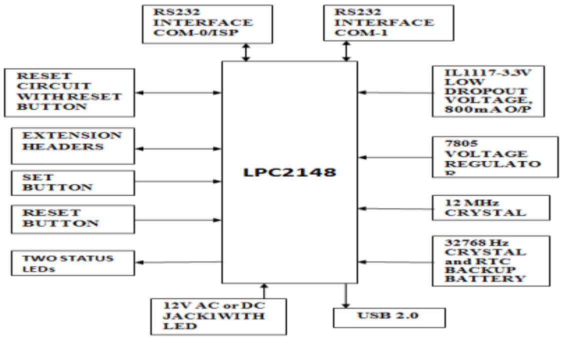

Figure 1: Highlights of ARM7 (LPC2138) microcontroller.

B. Serial Communication With Keyboard

This project uses PS/2 keyboard. Since most of the keyboards are USB keyboards which belonging to the modern era keyboard. It consists of all alphabets, numerical and numbers. When a key is pressed, the keyboard sends a break code. This is a simple implementation which ignores release keys for the most part. In this project only alphabets and numbers are used.

C. Interfacing The Lcd

A 16x2 Liquid Crystal Display can display 2 lines with 16 characters per line. Each character displayed is with 5x7 matrix.

This Liquid Crystal Display has 2 registers viz.

- Data Register DR – The displaying data is stored in this register. The data to be displayed is ASCII value of the character on the LCD.
- Command/Instruction Register C/I R- Which stores command instructions given to the LCD.

PIC16F877A

The family of PIC, PIC16F874A and PIC16F877A available with 44 packages can be used. 40 pin IC Peripheral Interface Controller 16F877A is suitably used for this project. This consists of 8 A/D input channels, 14 kB flash memory, 5 I/O ports with 15 interrupt handling capacity. Parallel Slave Port is available only on the package 40 or 44-pin devices.

D. Wireless Transmission Of Data

For the wireless transmission purpose CC2500 RF Module is used. This is a trans-receiver module that provides an easy RF communication at 2.4 GHz. Data can be transmitted or received for CC2500 RF module by any standard TTL source.

CC2500 RF Module supports multiple 9600 bps baud rates. It works on the ISM (Industrial Scientific and Medical radio) band 2.4 GHz. No complex wireless connection software or intimate knowledge of RF is required to connect serial devices. Designed to be as easy to use as cables and no external antenna is required. It can be used as a plug and play device which is powered with 5V DC supply.

E. Software Requirements

The algorithms are coded and developed in C using Keil and finally implemented on commercially available embedded microcontroller. C is a computer programming language used for general purpose and it supports structured programming. C can comfortably construct to machine instruction. C is used to build many apps and Operating System.

Proteus stands for Processor for text easy to use. It is a fully functional and procedural programming language.
Proteus supports languages like C, BASIC, Assembly, clipper etc. This project is programmed with C. The Keil Development tool is used. Keil is used for ARM controller and LPC 2XXX series.

IV. SYSTEM IMPLEMENTATION

As explained in section II, this project is divided into two hardware units - transmitter and receiver. This section explains the composition and functional operation of the transmitter block.

The block diagram of the proposed transmitter system is shown in Figure 2. In this project, ARM7 LPC2148 processor is used, this processor controls all the functions like data capture, storing and data analysis. PS/2 keyboard is used as an input device to send the input message. PIC16F877A is used as the interface between keyboard and ARM controller. CC2500 RF is used as the transmitter, which transmits message over air. For display purpose 16x2 lines LCD is used.

At the receiver side, we need a system with accuracy necessary for capturing data which is transmitted over a distance. The block diagram in Figure 3 shows the proposed receiver system.

The receiver consists of the same CC2500 RF module. The receiver uses another ARM LPC2148 for storing the received data. As shown in Figure 3 ARM processor receives the data from interfaced RF module.

Figure 2: Modules used for transmission of message.

We need a dictionary which is stored with large number of words. The TTS256 plays the role of this, which is an 8-bit microprocessor programmed with 600 letter-to-sound rules. It is real-time translation of English text to allophone addresses. Combined with TTS SpeakJet it provides a complete text to speech solution for outcome.

Figure 3: Receiver block with interconnection between modules.

V. RESULT AND DISCUSSION

This section presents the execution procedure with the results and discussion of these results as follows.The performance of the implemented system is evaluated. Whenever a disabled person wants to communicate with other persons, he will use keyboard to share his thoughts. As a key is pressed the entered key will display on transmitter LCD, if the key pressed is not displayed, then there will be LED indication, which tells whether the entered character or number is transferred or not properly. In case of improper transfer, it retried until successful transmission. After successfully entering twenty characters, the sending message will reach the transmitter LCD. ARM controller keeps track of buffer size. Once it reaches cutoff it will start sending messages to receivers.

With Xbee transmitter we can now transmit the data up to 50meters without any errors. The sending indications will be displayed in Xbee module. At the same time receiver Xbee will indicate that, typed message is being received. Before all this, the receiver setup should be on, which is confirmed by a ready signal. As the receiver module captures the data, it will store in the ARM buffer. Depending on the end of the message, the ARM will send data to receiver LCD and to the speaker via TTS-SpeakJet. Finally, we can able to receive the message which the disabled person entered at a distance of 50 meters away successfully.

Execution of the system is as follows:

Figure 4 shows the transmitter module, which consist of standard keyboard, peripheral interface controller 16F874A, ARM controller, LCD display and the Xbee. The keyboard, which will send pressed key data to the interfaced PIC via PS2 connector. There are LED indications on the PIC, Xbee and on ARM board. The LED glows at the transmitter module indicates the entered data is transferring from keyboard to an ARM.

International Journal of Research (IJR)
e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015
Available at http://internationaljournalofresearch.org

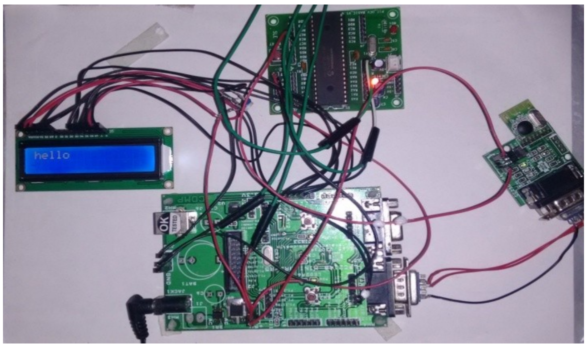

Figure 4: Transmitter Module

As the transmitter module and receiver module are powered up with DC, the default welcome message will pop on the screen, which displays the message “RealTime-TTS LPC2148”. The display contrast can be adjusted according to visibility.

As the person begins to press keys, the appropriate typed key will be displayed. The key pressed is evaluated and it is put into a loop for count. At each key press the character count is increased, after typing 16 characters, automatically the transmitter system initiates to send the data to Xbee transmitter. The sending indication will be given to the user.

The distance between transmitter and receiver is set up at 50 meters. Looking at the receiver side setup, the receiving indication will display as shown in Figure 5. Here the indication message “Real time-TTS Receiving” is displayed, and this will take place after the immediate transfer from transmitter module.

Figure 5 shows the complete receiver system, which consist of Xbee, ARM controller, LCD, TTS, SpeakJet and the speakers.

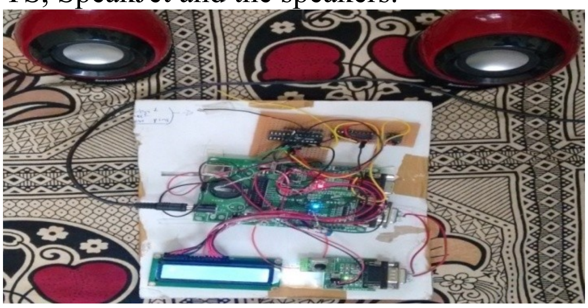
Figure 5: Reception Module.

As the receiver setup is powered, “ready” indication will be displayed on the LCD and simultaneously ready voice output will come which indicates the system is ready to receive, also confirms that all receiver setup is working well. Xbee which receives the data sent from the transmitter and the receiving is indicated by a green LED signal.

At the receiver, after the ready state each character starts displaying, simultaneously speech out from the speaker will output in a louder and clear voice.

VI. CONCLUSION

Embedded based implementation of real time text-to-voice conversion has been successfully accomplished. The proposed technology improves the communication gap between dumb and blind people.

The real time capability of transmission of text from the keyboard is buffered and displayed in the source or base station and simultaneously transmitted over air media. The system can deliver the message up to 50 meters. However, it can be extended up to 100 meters with the help of RF module. At the receiver side the system has the accuracy in receiving input message, which will be buffered and displayed at the receiver station. It can simultaneously provide clear voice and display output. People can easily understand the received voice message.

The hardware modules which are used in the proposed project are:

- Cost effective.
- Readily available in the market.

The system must be built in such a way that it can be made easily portable and used at any location. This can be accomplished by making a portable keyboard instead of using computer keyboard which consists of extra key words, symbols and functions which are not necessary for human communication. Therefore a keyboard with just alphabets, numbers, and few symbol keys can be used so that system can be made compact.

The whole setup can be done in such a way that hardware board used to interface keyboard and the display should be small as possible. The display size should vary according to environment need that is flexibility should be incorporated in display size. We can utilize this system in a conference or seminar halls. The typing and the output delivering can be made real time without any delay. The communication distance can be still increased with use of different transceivers. The speed of transmission and accuracy of receiving message plays the main role for proper communication and hence there need to concentrate more in this direction.

ACKNOWLEDGMENT

I express my sincere thanks to V. Nattarasu, Associate Professor, Department of Electronics and Communication, SJCE, Mysuru for his constant and valuable guidance in completing this project.

REFERENCES

[1] ARM based implementation of Text-To-Speech (TTS) for real time Embedded System Abdul Rawoof, Kulesh, Kailash Chandra Ray Dept. Of Electrical Engineering, Indian Institute of Technology Patna, India-800013.

[2] HaileMariam S. and Prahallad K., “Extraction of linguistic information with the aid of acoustic data to build speech systems,” in Proceedings of IEEE Int. Conf.Acoust., Speech, and Signal Processing, Honolulu, USA, 2007.

[3] T. Dutoit, An Introduction to Text-to-Speech Synthesis, Dordrecht/Boston/London: Kluwer Academic Publishers, 1997.

[4] Hunt A.J. and Black A.W., “Unit selection in a concatenative speech synthesis system for a large speech data base, in Proceedings of IEEE Int. Conf. Acoust., Speech, and Signal Processing, 1996, pp. 373–376.

[5] Research Article-microcontroller based gesture recognition System for the handicap people Bhavina Patel, Vandana Shah, Ravindra Kshirsagar.

[6] C. Bickley, A. Sydal, and J. Shroeter, ‘‘Speech Synthesis,’’ in The Acoustics of Speech Communication,J.M. Picket, Ed., Boston, NY: Allyn and Bacon, 1998.

[7] Recent U.S. Disability Statistics from the Census Bureau.

[8] James L. Flanagan, "Speech Analysis, Synthesis and Perception", Springer-Verlag, 1965, pp. 166-167.