

# Microcontroller Based Speech to Text Translation System

Uko Victor Sorochi, Gachada Benard Dubukumah, and Ibrahim Kafayat Ayosubomi

**Abstract**—For the purpose of an effective communication, speech becomes a convenient conduit to convey messages as an important activity in human life. The need for better reception of voice messages between humans and the environment they interact with becomes a trend and an area of interest that needs to be studied for improvements. A speech to text translation system is an embedded based design that convert analogue signals particularly voice from an input into digital signals that a computer or any electronic device can understand and perform a required task or display the equivalent digital signal in text on a screen. Speech translation systems mitigates the bottlenecks to an efficient communication caused by other varieties of communication methods. Even though speech translation and recognition designs haven't been well explored for electronic integration due to complexity and variation of sound signals from sources, this low cost, simple and portable project was incorporated to serve as a substratum and alternative in speech to text translation designs using microcontroller, in other to bridge the gaps in the world of human communications. This research paper addresses design methodology, limitations, recommendations and applications of the implemented speech to text translation system for improved communication reception.

**Index Terms**—Android Application, Embedded Design, Speech, Speech Translation and Text.

## I. INTRODUCTION

Every conception in the mind of a speaker is conveyed to audience through speech. The human to machine interface researches have been spurred by using speech to interact with computer and a wider audience on a faster and wider note. With modern processes, algorithms and methods, we can process speech signals easily into recognized text. [2]. Text can be displayed on screen for better reception of messages conveyed. This forms the aim of this project.

Presently, speech recognition and translation systems can analyze thousands of words from mainstream languages such as English, German, Spanish, Chinese, and Hindi etc. [3] The main requirement of every speech to text recognition and translation system is a database which will compare speech frequencies. This data base is already provided by Google Cloud for universal synchronization. In this project design, an application was developed for sending SMS messages based on Google Speech

Recognition Engine (Google Cloud). The application which runs on android devices allow users to input spoken information and send voice messages as desired text messages. The voice recognition and speech translation system is the devices capacity to understand spoken instructions and give equivalent text display on a screen.

An analogue to digital converter was used which convert varying analogue voice signals into digital signal pulses to be displayed on a screen. The brain of the project implementation was shouldered on a microcontroller called ATMEGA 328p. With the project implementation, the draw backs to efficient communication will be tackled to a minimum

## II. REVIEW OF KEY BUILD UPS TO THE PROJECT

### A. Liquid Crystal Display (LCD)

Liquid crystal display is an electronic display module that finds a wide range of application in circuits. It is preferred over seven segment display and matrix boards because of its portability, ease to interface with ATMEGA328p and its affordability. A 16 by 2 LCD used in this project means it can display sixteen character per line and there are two of such lines. There are two types of connections made when using the LCD. They include the eight-bit mode and the four-bit mode. When using the eight-bit mode, eleven out of the sixteen pinout of the LCD are used to transfer data per byte while the four bit uses seven out of the sixteen pinout to transfer data per byte. The four-bit mode was used in this project as it is much economical than the eight bit. [6].

For wider audience outreach, a wider screen was incorporated to the project using the RS232 cable

### B. Android Application

Android is a software environment for mobile devices that includes an operating system, middleware and key application. Main features of Android operating system are:

- Enables free download of development environment for application development.
- Free use and adaptation of operating system to manufacturers of mobile devices.
- Equality of basic core applications and additional applications in access to resources.
- Optimized use of memory and automatic control of applications which are being executed.
- Quick and easy development of applications using development tools and rich database of software libraries
- High quality of audiovisual content, it is possible to use vector graphics, and most audio and video format

---

Published on December 31, 2019.

U. V. Sorochi was with Department of Electrical/Electronic Engineering, Modibbo Adama University of Technology, Adamawa State, Nigeria. He is now with Department of Electrical/Electronic Engineering, Abubakar Tafawa Balewa University, Bauchi State Nigeria (e-mail: sorochivictor@gmail.com).

G. B. Dubukumah is with Department of Electrical/Electronic Engineering, Modibbo Adama University of Technology, Adamawa State Nigeria (e-mail: g.dubukumah@gmail.com).

I. K. Ayosubomi is with Department of Electrical/Electronic Engineering, Abubakar Tafawa Balewa University, Bauchi State Nigeria (e-mail: ibrahimkafayatubomi@gmail.com).

- Ability to test applications on most computing platforms, including Windows, Linux...

The Android operating system (OS) architecture is divided into 5 layers. The application layer of Android OS is visible to end user, and consists of user applications. The application layer includes basic applications which come with the operating system and applications which user subsequently takes. All applications written in the Java programming language. Framework is extensible set of software components used by all applications in the operating system. The next layer represents the libraries, written in the C and C++ programming languages, and OS accesses them via framework. Dalvik Virtual Machine (DVM), forms the main part of the executive system environment. Virtual machine is used to start the core libraries written in the Java programming language. Unlike Java's virtual machine, which is based on the stack, DVM bases on registry structure and it is intended for mobile devices. The last architecture layer of Android operating system is kernel based on Linux OS, which serves as a hardware abstraction layer. The main reasons for its use are memory management and processes, security model, network system and the constant development of systems. There are four basic components used in construction of applications: activity, intent, service and the content provider [5]

#### C. Microcontroller (ATMEGA328p)

This is a high-performance Atmel's 8-bit microcontroller. It is a low-power, high-performance based microcontroller which combines 32KB (kilobyte) ISP (in-system programming) flash memory with read-while-write capabilities, 1KB EEPROM (electrically erasable programmable read-only memory), 2KB SRAM (static random-access memory), 23 general purpose I/O (input/output) lines, 32 general purpose working registers, a single 16-bit Timer/Counter with independent prescaler, compare and capture modes. Other core features of the ATMEGA328 microcontroller are wide operating voltage range: 1.8V to 5.5V, maximum operating frequency of 20 MHz, Data retention for 20 years at 85°C and 100 years at 25°C and 10 bit, 6 channel analog to digital Converter. With these unique features, the ATMEGA328p was preferred for the project [6].

#### D. Bluetooth Module (HC-05)

It was used in the project to establish serial communication between the voice application and the microcontroller. It has two modes of operation namely: command mode and data transfer mode. In the command mode, it is connected as a master with the Arduino development board and then to the computer system to enable changes in device settings such as name of device, baud rate specification, status of device etc. in the Data transfer mode, it is connected in a slave configuration to a microcontroller. It receives Bluetooth signals from an external source, perform appropriate translation and pass the instruction to the microcontroller for action.

It has the following features:

- It is used for many applications like wireless headset, game controllers, wireless mouse, wireless keyboard and many more consumer applications.

- It has range up to <100m which depends upon transmitter and receiver, atmosphere, geographic & urban conditions.
- It is IEEE 802.15.1 standardized protocol, through which one can build wireless Personal Area Network (PAN). It uses frequency-hopping spread spectrum (FHSS) radio technology to send data over air.
- It uses serial communication to communicate with devices. It communicates with microcontroller using serial port (USART). [7].

#### E. Max 232

It is an integrated circuit (IC) embedded in a single chip and act as a voltage level converter. MAX 232 is capable of converting 5V TTL (Transistor Transistor Logic) level to TIA/EIA-232-F level and can take up to +30V input. It is normally used for the communication between microcontrollers and Laptop/PC. We can use MAX-232 to convert TTL voltage level to RS232, (Recommended Standard 232) and vice versa. [8].

#### F. Rs 232 Cable

In telecommunications, RS-232 is a standard originally introduced in 1960 for serial communication transmission of data. It formally defines signals connecting between a DTE (data terminal equipment) such as a computer terminal, and a DCE (data circuit-terminating equipment or data communication equipment), such as a modem. The standard defines the electrical characteristics and timing of signals, the meaning of signals, and the physical size and pinout of connectors. The current version of the standard is TIA-232-F Interface between data terminal equipment and data circuit-terminating equipment, employing serial binary data interchange, issued in 1997. The RS-232 standard had been commonly used in computer serial ports and is still widely used in industrial communication devices. [9].

#### G. Power Supply Unit

The power supply unit is the unit that energizes all other units of the system at a steady voltage. Functionally, the power supply converts A.C. voltage of 50 Hz power line to D.C voltage. In this design, 3.3V, 5V and 12V D.C voltages was required for various component making up the systems block.

The process for the conversion of A.C supply to D.C is termed rectification. There are two methods of rectification namely, full wave and half wave rectification. The full wave convert both circles of input AC voltage into pulsating DC voltage while the half wave convert only the positive half cycle of input AC voltage to pulsating DC voltage. In the power supply design, full wave rectification was used using a bridge rectifier because it offers better efficiency than the half wave rectification. [6]

### III. REVIEW OF RELATED WORKS

Pioneering research in the field of speech to text transformation started in AT&T Laboratories, Inc. Over the years, expanded researches for improvements have been made to the speech to text system for better service delivery. The continuous researches have produced facts and

recommendations as baselines for future enhancement of the system. Notable works in the speech to text translation system include:

The authors in [1] developed a speech to text system consisting of two components. First component was for processing acoustic signal which was captured by a microphone and second component was to interpret the processed signal, then mapping of the signals into words. Model for each letter was built using Hidden Markov Model (HMM). Feature extraction was done using Mel Frequency Capstral Coefficient (MFCC). Feature training of the datasheet was done using vector quantization (VQ) and feature testing of the data sheet was done using Viterbi algorithm. The main aim of the project was to recognize speech using MFCC and VQ technologies. The system was complex in operation because acoustic signals had to pass through series of rigorous processes before being outputted to text.

[2] Presented a survey that deals with different methods of speech to text conversion which was useful for different languages such as Phonem to Graphem method, conversion for Bengali language and HMM based speech synthesis methods. The objective of the paper was to recapitulate and match to different speech recognition systems as well as approaches for speech to text conversion and identifying research topics and applications. The speech to text system introduced was done via the internet by connecting to Google's server. The design loses its usefulness where there is poor internet access or none at all. Moreover, there was no backup system to augment for internet failure and fluctuations

In the work of [3] the accuracy and factors causing distortion in speech like noise and human traits of speaking was discussed. The paper also touched points like how error was found out or estimated with the help of algorithms like WAN (wide area network) and then refined using CRF (conditional random field) algorithms. With the aid of the error analysis, the output of future speech recognition systems will be more legitimate.

[4] Proposed an overview of general and specific techniques for better handling of variation sources in automatic speech recognition, mostly tackling the speech recognition system. The paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also insight of technique developed in each stage of speech recognition, while expressing best choices, relative merit and demerit. The major objective of the review paper was to summarize and compare different speech recognition systems and identify research topics which are at the forefront of the exciting and challenging field

[5] Developed a design similar to what was established in this paper using eclipse workbench. An application for sending SMS (short message service) which uses Googles speech recognition engine was also develop. The main goal of application voice SMS was to allow users to input spoken information and send voice messages as desired text messages. The speech recognition process which works over the internet was still very complex in operation.

#### IV. DESIGN METHODOLOGY

As an augment to the reviewed literatures, the project coordination was shouldered on a microcontroller called ATMEGA328p as aforementioned, in other to simplify the operation of the whole system's architecture. The design and construction of the implementation was segmented into two stages namely software and hardware divisions for ease of troubleshooting.

The hardware design was developed by integrating together the physical components making up the systems block diagram shown in Fig. 1. The software design was aimed at controlling the hardware components of the system. The software design was developed by interpreting the systems flow chat in form of programme codes written in the open-source Arduino environment. The implemented circuit diagram which shows how the physical component of the construction are electrically connected also forms a subsidiary of the hardware design.

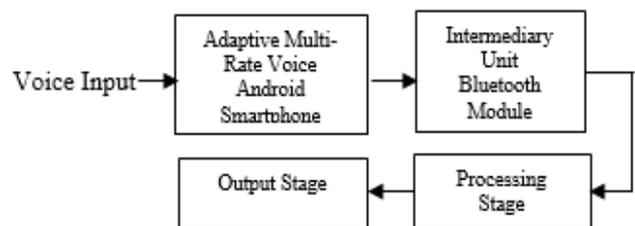


Fig. 1. Block diagram for the project construction

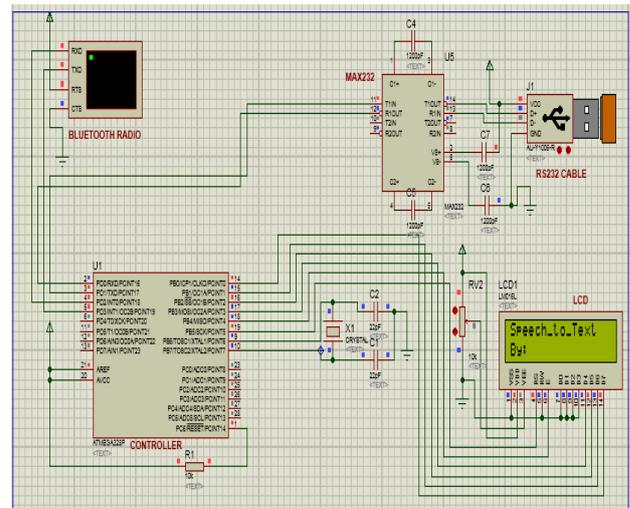


Fig. 2. Circuit diagram for the project construction

##### A. Circuit Diagram Description (Hardware Design)

When a speech input was made via an android phone's microphone, two applications were necessary for the operation of voice command recognition from such android device. An AMR (adaptive multi-rate audio) application which was the main application through which reception of voice commands from user was obtained and further Bluetooth synchronism was achieved by the Bluetooth module. Secondly, Google Voice Search application which formed the supporting application having an English library (or any chosen speech language) words by which a given command was understood and also translated into a string of characters for transmission to the Bluetooth module.

The module interfaced the given characters to the microcontroller by serial communication. The string of

characters was converted into digital text to be displayed based on programmed codes of the microcontroller. The microcontroller utilizes TTL for its operations. The Max 232 converts the 5V TTL level from the microcontroller to TIA/EIA-232-F level for interface to a data terminating equipment such as a computer screen (larger screen) using the RS 232 cable.

For personalized view of text, a 16 by 2 liquid crystal display (LCD) was incorporated

### B. Software Design

A flowchart was developed based on the hardware description. The flowchart formed the base for developing the programme codes in the open-source Arduino environment. The programme codes or instructions was used to control the hardware components of the project

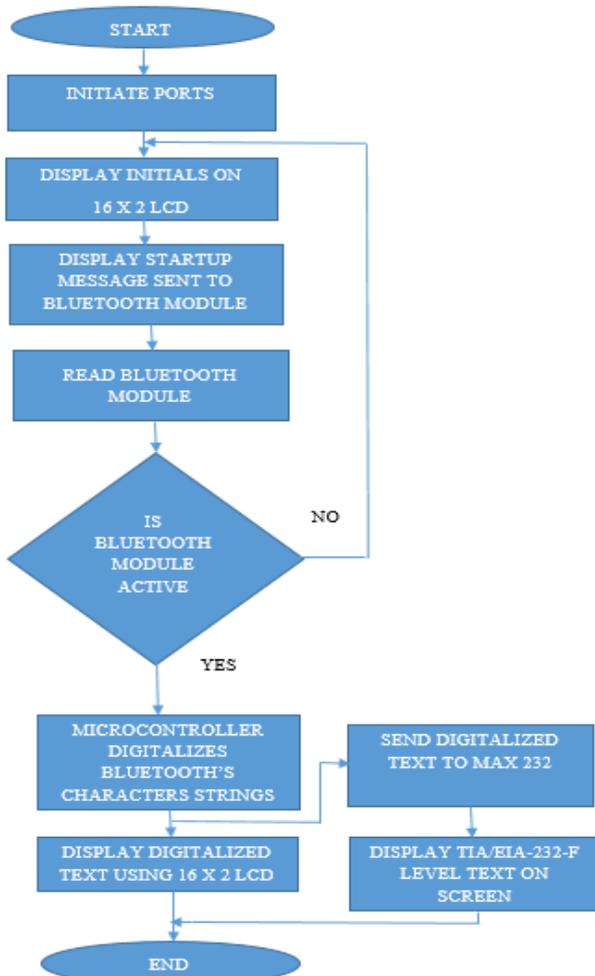


Fig. 2. Flow chart for the programme design

### C. Power Supply Design

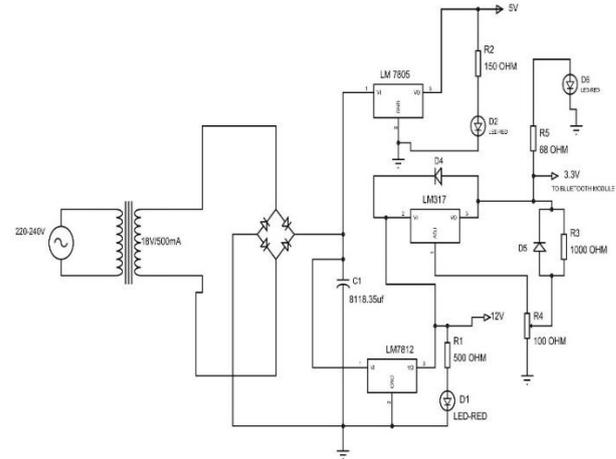


Fig. 3. Power supply circuitry

### D. Shunt Capacitor C1 Numerical Design

$$\text{The peak voltage} = V_{peak} = V_s \times \sqrt{2} \quad (1)$$

The peak voltage is the difference between the positive peak voltage waveform and the negative peak voltage waveform of the transformer secondary ac signal.

$V_s = 18V$  (Rating of the secondary voltage of the transformer used):

$$V_{peak} = \sqrt{2} \times 18 = 25.46V$$

$$\text{value of the rectified voltage} = V_{dc} = \frac{2}{\pi} V_{peak} \quad (2)$$

The dc value  $V_{dc}$  of the rectified voltage was calculated using (2):

$$V_{dc} = \frac{2}{\pi} \times 25.46 = 16.21V$$

Maximum load current is given by

$$I_m = \sqrt{2} \times I_{R.M.S} \quad (3)$$

The maximum load current is the maximum current the loads in the implementation can actually demand from the system's power supply unit.

$I_{R.M.S} = 500mA = 0.5A$  (Rating of the secondary current of the transformer used)

The maximum current the circuit loads can draw from the power supply was calculated using (3):

$$I_m = \sqrt{2} \times 0.5 = 0.71A$$

$$\text{Average load current} = I_{dc} = \frac{2}{\pi} I_m \quad (4)$$

The average load current was calculated using (4)

$$I_{dc} = \frac{2}{\pi} \times 0.71 = 0.45A$$

It was preferable to use a filtering capacitor that will hold the peak ripple voltage ( $V_{ripple}$ ) at approximately 1%. The ripple factor (RF) was important in deciding the effectiveness of the rectifier. That is, it indicated how much

rejection the rectifier had to have in other to achieve a certain noise level on the output

$$\text{Peak ripple voltage} = V_{\text{ripple}} = RF \times V_{dc} \quad (5)$$

$RF = \text{Ripple factor} = 1\% = 0.01$

The peak ripple voltage was calculated using (5):

$$\text{Peak ripple voltage } (V_{\text{ripple}}) = 0.01 \times 16.21 = 0.16$$

$$C_1 = \frac{I_{dc}}{F \times V_{\text{ripple}} \times 4\sqrt{3}} \quad (6)$$

$F = \text{Frequency of the source alternating current} = 50\text{Hz}$

The shunt capacitor  $C_1$  value was calculated using (6)

$$C_1 = \frac{0.45}{50 \times 0.16 \times 4\sqrt{3}} = \frac{0.45}{55.43} = 0.0081183F$$

$$C_1 = 8118.35\mu F$$

From the E series for capacitors available for commercial purchase, the rated capacitance for  $C_1$  was not available. Hence, available capacitors were parallel at minimum to make up for the rating of  $C_1$ . The higher the paralleled capacitance above the value of  $C_1$  the better the filtration of the resultant DC voltage.

Diode D5 and D4 in Fig. 3, was used as protection diode for the LM317 while resistors R4 and R3 was suitably chosen to account for 3.3V from LM317.

#### E. Resistors R1, R2 and R5 Numerical Design

Applying Kirchhoff Voltage Law through  $R_1$ ,  $R_2$  and  $R_5$ , with respect to their source voltages and there connected LED's gives (7)

$$V = V_D + I_D R_n \quad (7)$$

Where  $n = 1, 2$  and  $3$

$V = \text{The source voltage for resistor } R_n$

$V_D = \text{The voltage consumed by the LED from its data sheet specification} = 2V$

$I_D = \text{Current consumed by the LED from its data sheet specification} = 20mA$

Refer to (7),  $R_n$  was made subject of the equation and the result displayed as (8)

$$R_n = \frac{V - V_D}{I_D} \quad (8)$$

Using (8), at  $n = 1$  and  $V = 12$

$$R_1 = \frac{12 - 2}{20mA} = \frac{10}{20} = 500 \text{ ohms}$$

Using (8), at  $n = 2$  and  $V = 5$

$$R_2 = \frac{5 - 2}{20mA} = \frac{3}{20} = 150 \text{ ohms}$$

Using (8), at  $n = 3$  and  $V = 3.3$

$$R_5 = \frac{V - V_D}{I_D} = \frac{3.3 - 2}{20mA} = \frac{1.3}{20} = 65 \text{ ohms}$$

Since a resistor of 65 ohms was not available in the E Series for commercially available resistors, a 68-ohm resistor was used in its stead.

## V. TEST RESULT

The result expected from the project was obtained as

shown in Fig. 4 and Fig. 5. The computer system was able to translate all what was said through the microphone of the Android device. There was specification on the language spoken. Only English language was used as the medium of communication in this project because only English language library was installed in the android device from goggle cloud. Other languages libraries can also be downloaded based on preference. The goggle cloud has more than one hundred languages libraries available for users.

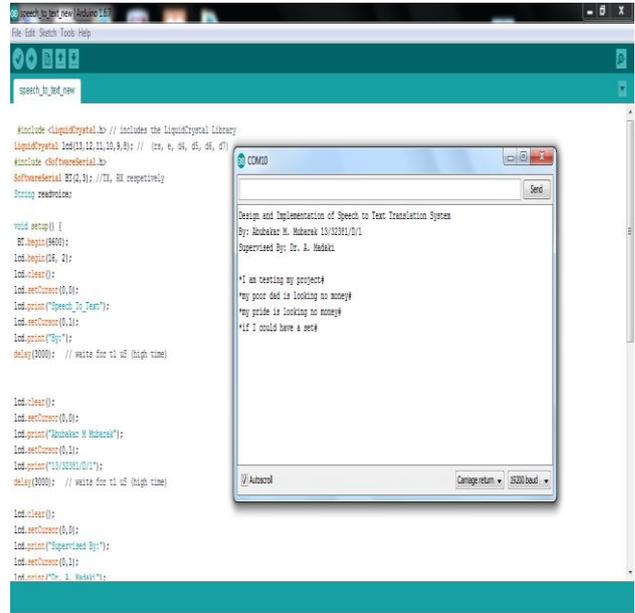


Fig. 4. Text display on open source Arduino Environment

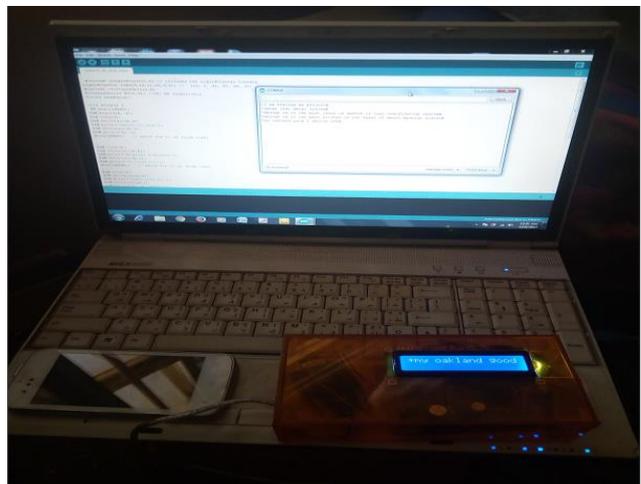


Fig. 5. Speech translated to text on screen and project's LCD

#### A. Limitations of the Project

- When a user goes outside the Bluetooth coverage range using the android device, speech translation becomes difficult.
- Noise interferences affects the genuiness of the output text
- There was absence of storage system for storing text displays for future use
- Only speeches with clear English language ascent was translated to text
- The project was only limited to communication with English language

## B. Recommendations

- Advancement should be made on the Bluetooth coverage range for ease of flexibility in the project usage
- A noise filter can be incorporated to prevent environmental noise interferences
- For future reference to displayed text, a storage system for the text messages should be incorporated
- Advancement should be made to incorporate multiple language libraries to the project to accommodate different users who can only communicate with their peculiar languages

## C. Applications of Speech to Text Translation System

- Speech to text translation system provides an alternative for surfing through the internet easily
- The system provides a revamped method of communicating with electronic gadget like automatic teller machines (ATM), multimedia devices, robots etc.
- The project will find application in home automation and security systems
- Speech to text system substitute the use of keyboard in communication
- The system will assist learned but handicapped people such as the deaf and dumb to integrate well in the society

## VI. CONCLUSION

The design and construction of a microcontroller based speech to text system has been presented in this research work. The system boast of its simplicity of usage, portability and less construction cost. The aim of the project which was to translate spoken words to text displays was achieved with proofs. The simplicity of the system's design and operation was achieved by integrating a microcontroller and segmenting the design phases into software and hardware construction for ease in troubleshooting the system.

With the development of mobile application for voice models, speech recognition will better enhance human to computer communications while giving possibility to manage mobile devices without installing complex software delegated for speech processing. Hence, aiding memory saving in mobile devices.

The limitations and recommendation for future advancement of the project have been highlighted. As a design alternative to reviewed literatures, this project will spice up future developments in the field of speech to text translation systems

## ACKNOWLEDGMENT

To whom much is given, much is expected. This necessitated the research team to express their unalloyed gratitude to Almighty God for His embellished enablement towards the success of the research work as an advancement to the body of knowledge.

It is creditable at this juncture to appreciate management and staff of the Department of Electrical/Electronic Engineering of both Modibbo Adama University of Technology Yola and that of Abubakar Tafawa Balewa

University Bauchi, Nigeria, for their warm reception and open arms during the project buildups.

As true as the saying goes that "Rome was not built in a day", lay emphasis on the fact that this project did not just come out of the blues. It rested on the building blocks laid by predecessors of similar project design, for whom the project team doff her hat in appreciation to all their inspirations in prints.

## REFERENCES

- [1] P. Das, K. Acharjee, P. Das and V. Prasad, "Voice recognition system: speech-to-text" *Journal of applied and fundamental science* vol. 1, pp. 191-195, November 2015.
- [2] P. Khilari, and V. J. Bhope, "A review on speech to text conversion methods" *International journal of advanced research in computer engineering & technology*, vol. 4, no. 7, pp. 3067-3072, July 2015
- [3] P. Padhye, and V. Chavan, and S. Mane M. "Speech recognition operation with efficient accuracy rate and factors affecting it" *International journal of advanced research in computer and communication engineering*, vol. 5, no. 9, pp. 767-782, September 2016.
- [4] S. Das, "Speech recognition technique: a review" *International Journal of engineering research and applications*, vol. 2, no. 3, pp. 2071-2087.
- [5] B. R. Reddy and E. Mahender, "Speech to text conversion using android platform" *International journal of engineering research and applications*, vol. 3, no. 1, pp. 253-258.
- [6] G. B. Dubukumah, U. V. Sorochi and A. Salisu "Fire monitoring, prevention and control system for market Shops" *Communications on applied electronics*, vol. 7, no. 31, pp. 26-31.
- [7] Bluetooth Module HC-05. (n.d). Retrieved from <https://www.electronicwings.com/sensors-modules/bluetooth-module-hc-05->
- [8] Introduction to MAX 232. (n.d). Retrieved from <https://www.theengineeringprojects.com/2017/07/introduction-to-max232.html>
- [9] About RS 232. (n.d). Retrieved from <https://en.wikipedia.org/wiki/RS-232>.



**Uko Victor Sorochi** obtained his B.Eng. in Electrical/Electronic Engineering from Abubakar Tafawa Balewa University Bauchi State, Nigeria in 2018. His research interest are in embedded systems, amplifier designs and renewable energy systems  
Mr. Sorochi is an associate member of Environment and Safety Management Institute and Chartered Institute of Contract Project and Facility Management



**Ibrahim Kafayat Ayosubomi** obtained her B.Eng. in Electrical/Electronics Engineering from Abubakar Tafawa Balewa University Bauchi State, Nigeria in 2017. Her research interest are in electrical machines, power and embedded systems.  
Miss Ayosubomi is currently pursuing her MSc in Electrical/Electronic Engineering at Abubakar Tafawa Balewa University Bauchi.



**Gachada Benard Dubukumah** obtained his B.Eng. in Electrical/Electronic Engineering from Modibbo Adama University of Technology Yola, Nigeria in 2009. He received a M.Sc. degree in Electronic and Electrical Engineering from Loughborough University UK in 2017  
Mr. Dubukumah is currently a lecturer with the Department of Electrical/Electronic Engineering, Modibbo Adama University of Technology Yola, Nigeria. His research interest are in power and

embedded systems

Mr. Dubukumah is a Member of the Nigeria Society of Engineers (NSE), and a registered member of the Council for the Regulation of Engineering in Nigeria (COREN)