International Journal of Advanced Science and Technology
Vol. 29, No. 12s, (2020), pp. 1295-1301

Real-Time Detection Of Human Speech Emotion Using ATMEGA

[1*]Hemanta Kumar Palo, [2]Niharika Pattanaik, [3]Badri Narayan Sahu
Department of ECE, ITER, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar Odisha, India
[*]hemantapalo@soa.ac.in

Abstract

Emotion detection is a process of automatically recognizing the emotion of a speaker based on features. A suitable feature extraction technique can make it possible to detect and inform an individual about his/her emotions. It can be used for safety and security by alarming an individual of an angry mob or by informing him/ her about the voice tempo during a public speech or presentation. The goal of this work is to build simple yet complete and representative automatic speech emotion recognition system. Due to a limited availability of voice sensors, this work tests the emotional voice samples using the amplitude, energy and pitch features. However, one can have a large amount of database and different sensors to develop a for more accurate system.
Keywords*: Emotion, Detection, amplitude, pitch, accuracy*

1. Introduction

Emotion detection is currently found to be an important and interesting part of speech analysis. The analysis can be done by selection of an effective parameter or by combination of several parameters to achieve the desired level of accuracy. In this regard, prosodic features such as the Energy, MFCCs, pitch, timbre, vocal tract frequencies, etc. are found to be effective introduced in many literatures [1]-[5]. Advanced research in emotional encounters will be based on the growing understanding of the definition of emotions, reliable feature extraction, and efficient classifiers. Software specified for human-robot interaction needs to receive human emotions and optimize their behavior [2].

The most natural way of recognizing basic emotions in humans is the extraction of a set of reliable features from speech signal or facial expressions. A numerous study has been conducted during last few decades aiming to improve the robustness and reliability of the human emotion recognition from voice samples [6]-[9]. These studies have attempted to investigate the appropriateness of using acoustic (pitch, intensity and formant), phonetic properties of emotive speech with the freeware program. A system for detecting and classifying emotions based on speech analysis can be considered as knowledge-based systems designed to analyze speech (words and sentences). Some works tried to incorporate feature extraction and analysis the different aspects of emotions in speech signals [10]-[11]. Although, affective high-level human interaction with robots is still far behind. The objective of this work is to detect human speech emotion economically as it does not include by not including any expensive tools. Instead, it uses Atmega16 microcontroller as the base unit for detection of the emotions in real-time. It's a onetime investment and will yield results for a longer period. With proper advancements made in this prototype it can easily detect and can also give suitable information to the users by using other features such as words per second etc. It can be helpful to the employees of the sales department and help them to effectively perform their tasks. It is a wireless device and only takes sound as input so is free from any complications. Further, it is easier to repair them by using a wide range of database.

2. Experimental Set-up

The work deals with the emotion detection technique based on the pitch and energy of the sound source efficiently in a MATLAB environment. It also has wide range of engineering applications such as

creating an interface with the microcontroller or sound source as sound can serve to analyze the voice setting. The project deals with the issue as the work is divided into two parts:

a) estimation of pitch and energy of sound signal
b) Interfacing the microcontroller with LCD and LED's

The steps of real-time speech emotion detection are:

- To detect sound and process it: The incoming sound signal is passed through the microphone.
- Feature Extraction: The sound intensity, pitch and energy is extracted from the speech samples.
- Microcontroller programming: Microcontroller is programmed through VLR studio and it is interfaced with the MATLAB using Bluetooth device.
- Interfacing: Interfacing is done using universal synchronous asynchronous receiver transmitter

The block diagram of the emotion detection system is shown in Figure 1. The emotional speech voice is recorded in a mobile set (Nokia set) in MP3 format. The signal is fed to the microcontroller through the Bluetooth device. The program is run for the chosen features extracted. A threshold is set for different speech emotions based on the magnitude of the feature which is indicated by the display devices such as LED and LCD.

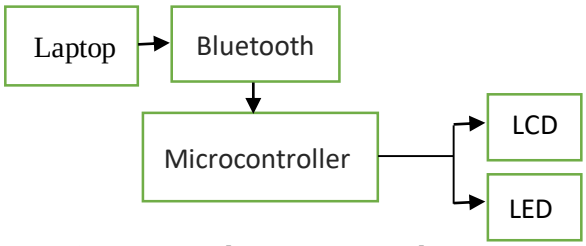

Figure 1: Real-time Emotion detection system

3. Technical Specification

ATMEGA16 description

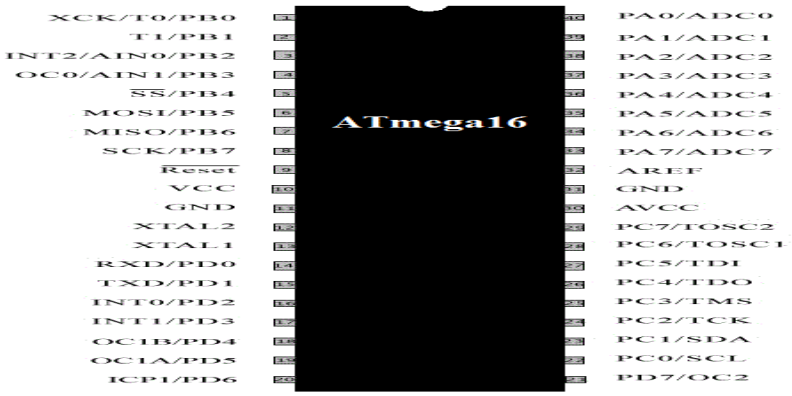

Figure 2: TMEGA-16 Pin configuration

ATmega16 is an 8-bit 40 pin microcontroller having the pin configuration as shown in Figure 2. It is a high-performance microcontroller of Atmels Mega. It consumes less power and works at a maximum frequency of 16MHz with a 16kb of flash memory. There are 32-bit input and output lines which are divided into 8-bit ports namely port A, port B, port C, port D.

IC 7805 description

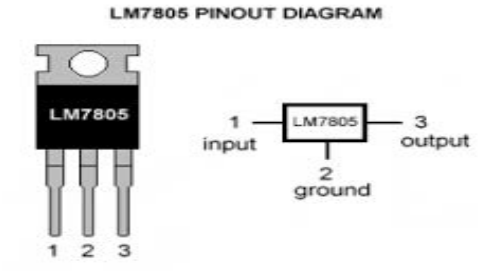

Figure 3: Pin diagram of LM7805

IC 7508 shown in Figure 3 is a member of 78xx series of voltage regulator IC's. The xx is the fixed value of voltage it is designed to provide. The conventional voltage sources may not give a fixed range of voltage outputs because of its fluctuations. Therefore 7805 IC is used to maintain a constant voltage value. Capacitors can be connected at input or output with proper values.

LCD Display Unit

The LCD unit used in this work is to indicate the desired emotion based on the extracted feature value. It displays the desired emotion when fed with the emotional dependent features and is shown in Figure 4.

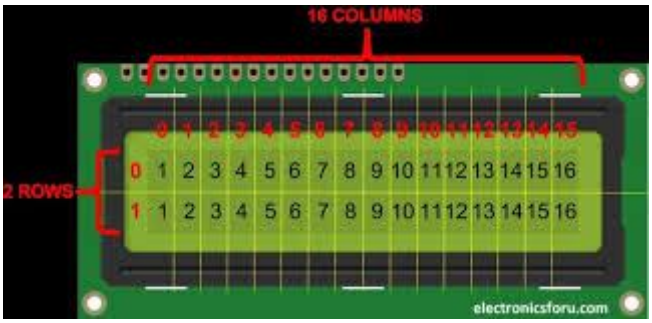

Figure 4: LCD Unit

This LCD has 2 registers command and Data. The command register is selected when signal is low and data register when signal is high.

LED Display Unit

The LED unit used in this work is shown in Figure 5.

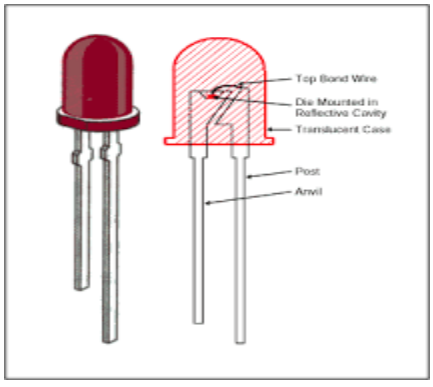

Figure 5: LED unit

It is used to display the desired emotion based on the emotional dependent feature value. Light emitting diode is a p-n junction semiconductor light source. We can adjust the brightness of the LED using a potentiometer.

Bluetooth Module (HC-05)

The Bluetooth module is shown in Figure 6 and is mainly used for short range wireless connectivity. It is an easiest and fastest way to make a connection. It uses short wavelength of ultra-high frequency radio waves with a range of approximately 10 meters. Before pairing it has an LED which blinks twice in a second and after the pairing the blinking reduces to once in every 3 sec.

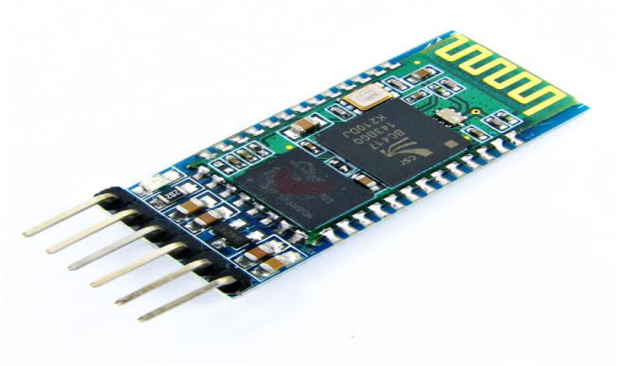

Figure 6: The blue tooth module

3. Discussion

The MATLAB platform has been used to extract the desired pitch, magnitude and energy of an emotional speech signal. A database has been developed for 3 main emotions namely anger, sad, and happy states. The AVR studio software developed by ATMEL has been used to program the microcontroller Atmega16. Program for interfacing of LED and LCD was done in AVR studio and its path was copied and transferred into the microcontroller. The AVR studio has its Editor, simulator and programmer. It has an integrated C compiler GCC (GNU C compiler) and gives a scope for developing tools like Q-Touch Studio.
Hardware implementation was done on a Vero board and the interfacing was done via Bluetooth. Power supply was taken from the laptop because it gives a 5v output which we need for this experiment. It is also possible to use power supply an adapter by connecting a 7805 transistor. There are three LEDs connected with a potentiometer to adjust the contrast of LED. A 16×2 alphanumeric LCD is used to display the status of the desired emotion. AVR studio was used to programmer the microcontroller and through the microcontroller message were delivered to LCD and LED.

The overall design of the system is shown in Figure 7.

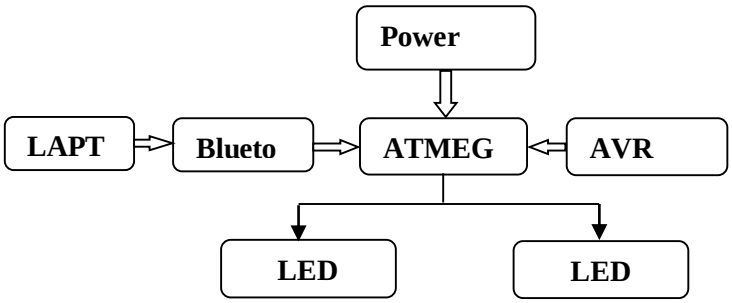

Figure 7: The overall system design

The steps of operation of the emotional speech recognition system are as follows:

- The system is turned on with a 5v supplies power to the micro controller board from the laptop.
- The MATLAB data is sent to the microcontroller via Bluetooth named HC-05.
- The Microcontroller is connected with LCD at port C and LED data to Port B of the microcontroller. The microcontroller contains the program logic and acts like a circuit that holds the pulse until the next pulse is received.
- The LCD uses a 12*2 display to print the message signal. The data is sent through USART (universal synchronous asynchronous receiver transmitter).
- Initially, the MATLAB code is generated for Bluetooth connection. The device name and type are given and the conditions are set to pair the device with the laptop. On successful connection of Bluetooth device with the computer, the microcontroller can used for sending data to LCD. The features of the speech signal are extracted using the MATLAB.
- The Hardware connections has been made on the Vero board and all the equipment were properly soldered. LCD was connected to port C of Atmega 16 microcontroller. LED was connected to port D. The 10 is used for 5V VCC supply and port 11 is grounded.

The total system was divided into sub modules or sub functions – power supply module, Microcontroller module, processing module and Testing module. The AVR studio was used to program the microcontroller. In the AVR studio, the interfacing of LCD and LED were coded and it was connected with the ports C and D of the microcontroller. The basic function of the microcontroller was to just send the data into LCD and generate a message in it. The processing module consisted of the MATLAB code for feature extraction of a signal. After coding the address of the speech signals were copied from the data base and program was executed. Based on various results the absolute value was found and accordingly conditions were fixed for detection of emotions. Hardware connections were made on the Vero board. Atmega 16 microcontroller was used and interfaced with 16*2 LCD and LED of 3mm, 1.5v. There are three LEDs for three different chosen emotions. ALR studio software was used for burning the entire program in the microcontroller chip.

3. Testing, Analysis and Evaluation

Two types of testing has been undergone:

1. Module Testing
2. Integration Testing

Module Testing: Module testing was done for different modules like microcontroller unit, power supply unit, LCD, LED module. Expected outputs were obtained against given inputs.
AVR studio-> Interfacing was done using software called AVR studio. Hex file was copied of the program and pasted into the program file of the microcontroller. Then messages were send using the microcontroller to the LCD and it was tested.

- **Bluetooth Module:** Bluetooth module was connected to the Laptop and it was turned ON. After some time (approx 1 minute) and the blinking rate of 'status LED' or 'network LED' was tested. (Bluetooth module will take some time to establish connection with Computer network). Once the connection is established successfully, the status/network LED blinks continuously every 3 seconds. This concluded that the Bluetooth module has successfully established network connection.
- **LED display:** To check whether the chip on the board is working properly or not the board was plugged into the USB port of a computer and the red LED power indicator on the board illuminates.
- **LCD:** After connecting the Atmega16 with the LCD and powering the driver by Laptop, the LCD showed message "Emotion Detection" and after the feeding the features of desired emotion, it showed "Emotion Detected" following the result namely 'ANGER' 'HAPPY' 'SAD'.

Safety and security are the two most important issues in today's world. This prototype attempts to detect the emotions accurately therefore the owner of this kit should first train the kit with his voice so that there is no scope of mismatch in the output. Since the instructions are being sent by the user through the Microphone there should be a proper noise filtering system in the kit to neglect the surrounding noise.

4. Conclusion

The work attempts to develop a prototype emotion speech recognition system using less number of hardware. It is a light weight system and easy to implement. However, the features used for detection of emotion are prosodic. More efficient utilization of reliable feature extraction techniques can improve the effectiveness of the system. Further, the system needs to be validated with other feature extraction techniques and with authenticated database.

References

[1] Darekar, R. V., & Dhande, A. P. (2016, February). Improving emotion detection with speech by enhanced approach. In *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 364-369). IEEE.
[2] Palo, H. K., & Mohanty, M. N. (2018). Wavelet based feature combination for recognition of emotions. *Ain Shams Engineering Journal*, *9*(4), 1799-1806.
[3] Rabiei, M., & Gasparetto, A. (2016, October). System and method for recognizing human emotion state based on analysis of speech and facial feature extraction; Applications to human-robot interaction. In *2016 4th International Conference on Robotics and Mechatronics (ICROM)* (pp. 266-271). IEEE.
[4] Mohanty, M. N., & Palo, H. K. (2020). Child emotion recognition using probabilistic neural network with effective features. *Measurement*, *152*, 107369.
[5] Stolar, M. N., Lech, M., Stolar, S. J., & Allen, N. B. (2018). Detection of Adolescent Depression from Speech Using Optimised Spectral Roll-Off Parameters. Biomedical Journal of Scientific & Technical Research, ISSN: 2574–1241, DOI: 10.26717/BJSTR.2018.05.001156.
[6] Palo, H. K., & Sarangi, L. (2020). Overview of Machine Learners in Classifying of Speech Signals. In *Handbook of Research on Emerging Trends and Applications of Machine Learning* (pp. 461-489). IGI Global.

[7] Deng, J., Frühholz, S., Zhang, Z., & Schuller, B. (2017). Recognizing emotions from whispered speech based on acoustic feature transfer learning. *IEEE Access*, *5*, 5235-5246.
[8] Palo, H. K., & Mohanty, M. N. (2020). Analysis of Speech Emotions Using Dynamics of Prosodic Parameters. In Cognitive Informatics and Soft Computing (pp. 333-340). Springer, Singapore.
[9] Joe, C. V., & Sugi, S. S. S. (2016). Optimal feature for emotion recognition from speech. *Afr J Basic Appl Sc*, *8*(3), 136-144.
[10] Palo, H. K., Behera, D., & Rout, B. C. (2020). Comparison of Classifiers for Speech Emotion Recognition (SER) with Discriminative Spectral Features. In *Advances in Intelligent Computing and Communication* (pp. 78-85). Springer, Singapore.Sato N, Obuchi Y (2007) Emotion recognition using Mel-frequency Cepstral coefficients. Inf Media Technol 2(3):835–848.
[11] Palo, H. K., & Behera, D. (2020). Analysis of Speaker's Age Using Clustering Approaches With Emotionally Dependent Speech Features. In *Critical Approaches to Information Retrieval Research* (pp. 172-197). IGI Global.