Journal of Information Storage and Processing Systems
UGC Care Listed
Volume 19, Issue 02, JUNE 2020.
Impact Factor: 6.303 ISSN:1099-8047

EMBEDDED SPEECH NORMALIZER

ROMA FATIMA1, AMTUL SALAM SOHA2, TAYYABA TASNEEM3, HAALA DEEBA ABBAS4, DR. SURYAKANTH V GANGASHETTY5, AND DR. M. A. NAYEEM6

1,2,3UG Scholar, B.E. Dept. of ECE, Deccan College of Engineering and Technology, Hyderabad - 500001, India.

4PG Scholar, MS, Speech Processing Laboratory, International Institute of Information Technology (IIIT), Hyderabad - 500032, India.

5Assistant Professor, Speech Processing Laboratory, International Institute of Information Technology (IIIT),
Hyderabad - 500032, India.

6Professor and HOD, Dept. of ECE, Deccan College of Engineering and Technology, Hyderabad - 500001, India.

1Email ID: romafatima51412@gmail.com 2Email ID: soha.ahmed8991@gmail.com 3Email ID: tayyabatasneen99@gmail.com

ABSTRACT:

Speech signals consist of a sequence of sounds and most of the time it is combined with noise and distortion. The noise and distortion cause the variations and reduce the intelligibility of the speech. Hence speech normalization was introduced for better communication. In today's technology, a pre-recorded audio waveform can be normalized. In this study, we discuss how to implement speech normalization in real time applications such as public oration and radio channel tuning. This is implemented through windowing the signal and gain updation method. The hardware used in the study comprises of the main STM32 circuitry containing electronic components such as, non- inverting Op-Amps and bridge rectifier. The STM32 takes care of gain updation method by constantly updating the

gain until the signal reaches the required voltage levels. Once the speech signal reaches the desired voltage levels, final normalized gain is obtained. The Speech Normalization is thus achieved.

Keywords: speech normalizer, gain updation, STM32, operational amplifier, bridge rectifier

1. INTRODUCTION

At a phonemic level, speech can be viewed as a sequence of basic sound units called phonemes. It is important to realise that phonemes are abstract linguistic units and may not be directly observed in the speech signal. The same phoneme may give rise to many different sounds or allophones at the acoustic level, depending on the phonemes which surround it. Also, different speakers producing the same string of phonemes convey the same information yet sound different as a result of differences in dialect and vocal tract length and shape.

Linguistic researchers study how a language changes by observing it. This is accomplished by looking at authentic data. For example, variation of speech is studied by looking at linguistic and social environments, then the data is analyzed as the change occurs. It is observed that the variation in speech is associated with the speaker's methodology, age, gender, geography and race. Although the listeners can perceive the speech in-spite of the variations of speakers, but there still exists some sort of noise and distortion [1].

Studies like [2] has made use of hybrid techniques such as Adaptive NeuroFuzzy Inference System (ANFIS) to remove noise from the degraded speech signal. Also, this study can- celled echo present in the speech signal by employing Finite Impulse Response (FIR) based Normalized Least Mean Square (NLMS) algorithm. Another variant of LMS approach called as Fractional

Least Mean square (FLMS) algorithm based echo canceller was used in [3]. Also Linear Prediction based methods have been used in the past to obtain dereverberant high quality speech. Examples of such studies include work done in [4]. Studies in [5] focussed on using weighted prediction error (WPE) method, along with spectral variance of early speech to produce dereverberation algorithm. This produced speech of high quality, besides reducing complexity of the system performance. But such linear prediction based techniques pose the problem of the linear prediction coefficients not getting properly optimised, therefore, few filters such as Kalman smoother was used in [6], that offered high robustness and were yet very resistive to the background noise and room reverberation present in the speech signal.

An embedded system is a combination of a computer processor, computer memory, and input/output peripheral devices, that has a dedicated function within a larger mechanical or electrical system. It is embedded as part of a complete device often including electrical or electronic hardware and mechan- ical parts. Because an embedded system typically controls physical operations of the machine that it is embedded within, it often has real-time computing constraints. Speech normalization refers to the reduction of variability or difference in the speech signals. This type of normalization deals with reducing the differences between pitch, loudness, frequency, etc., so that the signal reaches the desired level. For the present paper, this is achieved through the process of gain updation of the windowed signal and simulating the signal. This paper is organized as follows: Section 2 describes the components used and and why these components are chosen. Section 3 explains the proposed method and the circuitry of the speech

Journal of Information Storage and Processing Systems
UGC Care Listed
Volume 19, Issue 02, JUNE 2020.
Impact Factor: 6.303 ISSN:1099-8047

normalizer. Section 4 presents results and related discussion. Summary and conclusion are presented in Section 5.

II. COMPONENTS USED

A. Bridge Rectifier

A rectifier is a device that converts ac voltage to uni directional dc voltage. It decreases resistance to the current in one direction and increases resistance in the opposite direction [7]. A rectifier utilises unidirectional conduction devices like PN junction diode, Zener Diode, Schottky Diode, etc. Figure 1 shows Schottky Diode.

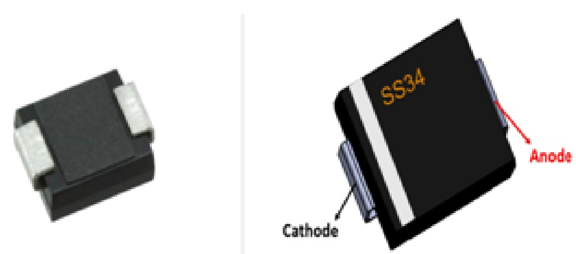

Fig. 1: Schottky Diode

A full wave bridge rectifier comprises of four diodes, eventually forming the four arms of the electrical bridge. These diodes in bridge circuit provides same polarity for both input and output [7]. In our project, we are using 4 schottky diodes as schottky diode has relatively faster switching speed and high efficiency.This schottky diode has a low voltage drop of about 0.5V and high forward current of 3 Amperes. These diodes have stable, near ideal I-V characteristics, fast switching time, low noise, and, most important of all, they can be made by a very simple process which is completely compatible with present-day integrated circuit technology [8].

The SS34 diode has a minimum voltage drop of around 0.2V across it when 0.1A is flowing through it, as the current increases the voltage drop across the diode also increases. The maximum current through the diode is 3A during when the voltage drop is only 0.5V and the maximum reverse voltage is 40V. It can also handle a maximum surge of 100A.

The main advantage of this circuit is it doesn't require a center tap on the secondary winding of the transformer. Hence, wherever possible, AC voltage can be directly applied to the bridge [7].

B. Operational Amplifier

An operational Amplifier is basically a differential amplifier whose output is proportional to the difference of two input's. In other words, Op amp has differential inputs and a single ended output where one input produces an inverted output signal, and other produces a non-inverting output signal.

1) ***Connection and details of Op Amp:*** An

Op amp has five connections, the inverting input terminal is supplied with negative voltage and the non inverting terminal is supplied with positive voltage. There is change in the sign at the output when the voltage is applied to inverting input, and there is no change in the sign at the output when the voltage is applied at the non-inverting input and it has two power supplies, positive and negative. Wide variety of useful circuits can be created using op-amp. Therefore, they are considered as general purpose building blocks [9].

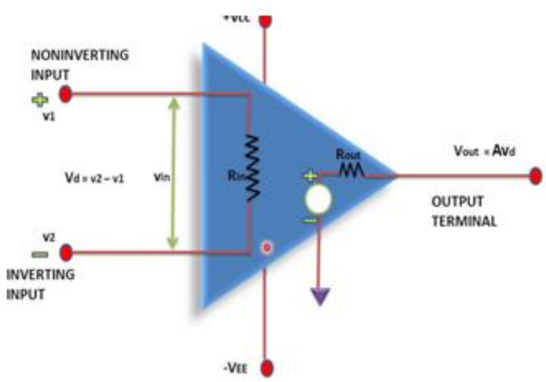

Fig. 2: Operational amplifier circuit.

2) ***Characteristics of ideal Op amp:*** Some of the characteristics of the Op-amp is that its differential voltage gain is very high i.e. equal to infinity. Input resistance R_i is equal to infinity, output resistance R_o is equal to 0, its characteristics do not drift with temperature and its bandwidth is infinity [10].

Fig. 3: TL081CDT circuit.

3) ***TL081CDT details and main circuit diagram explanation:*** Here we are using TL081CDT series of JFET transistors which have low input bias current which provide high quality and robustness. These are perfect companion to the micro-controller and analog signals and helps in enhancing the signal chain.

C. STM32

STM32 is a family of 32-bit microcontroller integrated circuits by STMicroelectronics. The STM32 chips are grouped into related series that are based around the same 32-bit ARM processor core. Internally, each microcontroller consists of the processor core, static RAM, flash memory, debugging interface, and various peripherals [11]. STM32 is preferred over other microcontrollers such as 8051, because it does not require external ADC and has bigger memory capacity and faster speed.

Fig. 4: STM32.

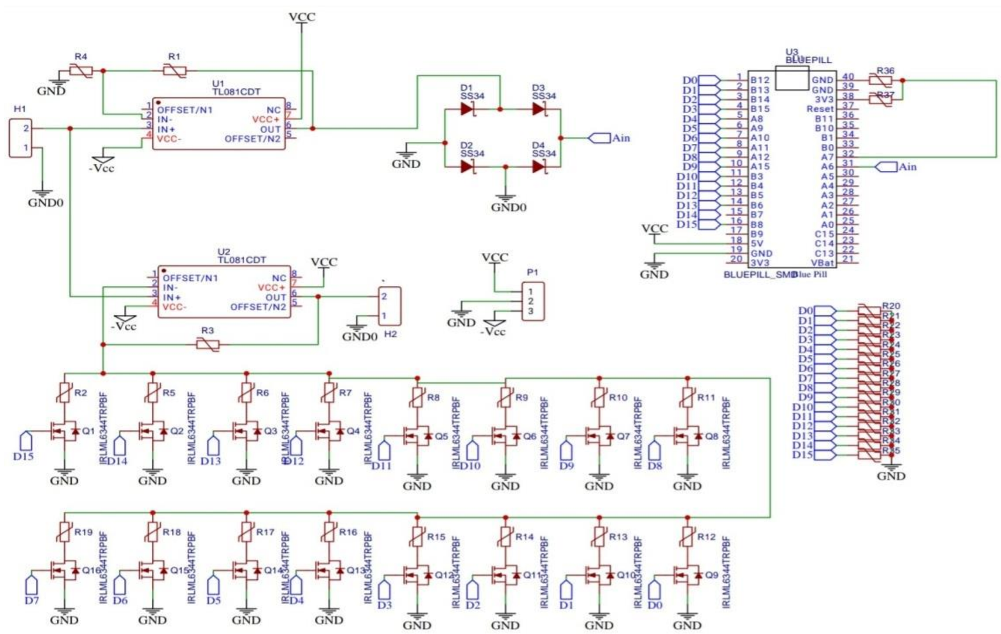

Fig. 5: Speech Normalizer Circuit Schematic Diagram

III PROPOSED METHOD

In our work, to implement speech normalization, first the signal is received through a microphone in a waveform. This analog double-sided signal is pre-amplified by passing it through two non-inverting amplifiers. During this process of pre-amplification, the gain value is changed to that of the amplifier and then the signal is passed on to the full-wave bridge rectifier, made up of 4 schottky diodes, which converts double-sided signal to single-sided signal. This

Journal of Information Storage and Processing Systems
UGC Care Listed
Volume 19, Issue 02, JUNE 2020.
Impact Factor: 6.303 ISSN:1099-8047

signal is then passed to a programmed STM32 MCU which windows the signal to find peak values and computes the gain required, then this gain is used to update the previous gain, thereby performing normalization. Finally, the signal reaches the required voltage levels, where it is given to a speaker as the output. This execution happens in real-time. The difference in signals can be observed by passing the input and output signals through a Cathode Ray Oscilloscope (CRO).

As shown in the main circuit diagram Fig. 5, two non- inverting JFET amplifiers U1 TL081CDT and U2 TL081CDT has been used. Through input header H1, the double sided analog signal is as input is passed to the non-inverting pre amplifier U1 TL081CDT as shown in the below Fig. 5. Here, the signal's initial gain configuration is changed by updating its gain value to that of the amplifier's one, thereby bringing the signal to sampling voltage levels. This is performed as the microcontroller can then sample this analog signal. Therefore, now the input analog signal is amplified through its gain updated with that of the amplifier.

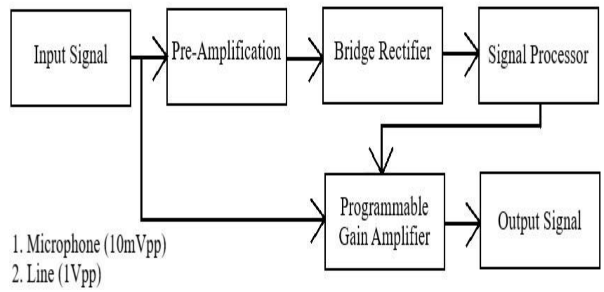

Fig. 6: Block Diagram of Speech Normalizer.

Also, the input signal can be passed through the second non-inverting amplifier U2 TL081CDT, that is then directly connected to the microcontroller part, that helps in attaining speech normalization.

The output of U1 is now passed through a full-wave bridge rectifier circuit, designed with 4 schottky diodes, so as to enable full-wave rectification. This rectifier takes the double sided analog signal and converts it into a single sided analog signal. The output of bridge rectifier circuit is now fed to this MCU through 'Ain'.

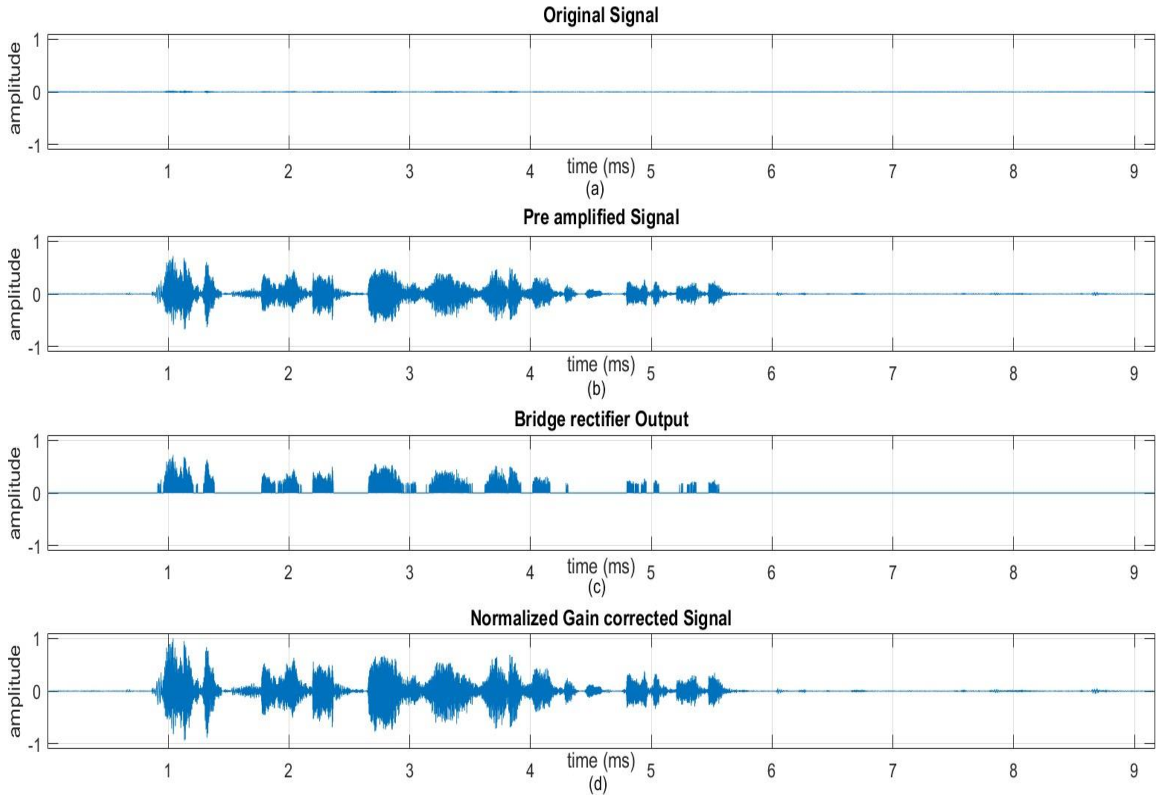

Fig. 7: Simulated Result (a) Original Unaltered Double Sided Speech Signal, (b) Original Signal After Gain Amplification, (c) Single Sided Signal After Rectification, (d) Final Normalized Gain Corrected Output Signal.

The purpose of STM32 is to properly update the gain and maintain its algorithm. The circuit can be operated in two configurations, when the signal reaches the MCU from bridge rectifier through U1 amplifier, and when the signal reaches the MCU directly through U2 amplifier. Both the configurations represented in, are used as per the application. The MCU has a group of resistors (R1, R2, R3 and so on), and transistors (D0, D1, D2 and so on). The resistors (R1, R2, R3 and so on) are designed in such a way that, the value of R2 is half of value of R5, R19 is half of R18, R6 is half of R7 and so on. Therefore, peak values are obtained to find the latest gain, and then this gain is used to update the previous gain, thereby performing normalization.

Hence, speech normalization is achieved.

IV RESULT AND DISCUSSIONS

In this section, the comprehensive experimentation results obtained by normalizing the input speech signal using gain amplification, rectification, crude filtering techniques. It is done by properly evaluating the value of the input signal and applying the above techniques to this input speech signal. The raw form of input signal (i.e., the original signal) tested during experiment is shown in Fig. 7(a) without any alterations.

Now the gain of the signal is amplified to the set value by paaing it through the amplifiers. One is fixed gain amplifier with a gain value of 8dB and the other is the variable gain am- plifier with a practical gain value of 9.46dB (theoretical value is 10dB). The resultant signal obtained after amplification is shown in Fig. 7(b). The amplified double sided signal is then converted into single sided signal using a bridge rectifier. The output obtained from the bridge rectifier is shown in Fig. 7(c) and the hardware implemented output is shown in Fig. 8.

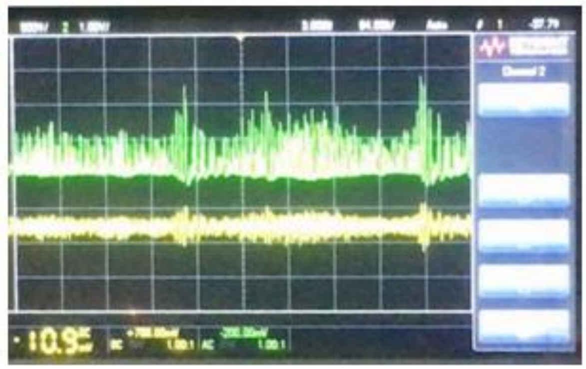

Fig. 8: Implemented Bridge Rectifier Output.

The final normalised output signal is shown in Fig. 7(d) and its implemented output is shown in Fig. 9. From the obtained resultant signal, it is clearly seen that this normalization technique produced the desired normalised gain corrected output signal, accurately.

Fig. 9: Implemented Final Normalised Output.

V SUMMARY AND CONCLUSION

This paper presents design analysis of Embedded speech normalization in real time applications. Embedded speech normalizer is created with the help of many hardware components such as Bridge rectifier, Operational amplifier and STM32 Microcontroller. This whole circuitry enables the signal to be windowed in order to obtain the peak values and compute the gain required. This gain gets updated in 256 steps and maximum gain is determined. This maximum gain obtained from the circuitry. Therefore, the speech signal is normalized whose maximum gain obtained is 9.46dB with each adjustment steps of 0.036. The simulation results obtained give the final normalized speech signal i.e., robust against the background noise.

VI ACKNOWLEDGEMENT

We would like to express our deepest appreciation a special gratitude to our Head Of Department (HOD), of Electronics and Communication Engineering (ECE), Dr. M. A. Nayeem for his contribution in simulating suggestions, counsel, support, advice and belief in us. Without his guidance and cooperation, this project would not have been possible. We would like to express our deep gratitude and appreciation to our external project guide Dr. Suryakanth V Gangashetty from Speech Processing Laboratory (SPL), at International Institute of Information Technology (IIIT), Hyderabad, for his support, valuable and constructive suggestions and care during the planning and development of our final year major project. We would also like to thank Ph.D. research scholar Mr. Likith Sai Dheekonda, for his advice and assistance for helping us in designing the circuit, fixing hardware and software problems. Our grateful thanks are also extended to MS research scholar Ms. Haala Deeba Abbas who helped, encouraged and coordinated us round the clock in writing this journal.

VII REFERENCES

[1] K. Johnson, “15 speaker normalization in speech perception,” *The handbook of speech perception*, p. 363, 2008.

[2] S. Dubey and A. Kolhe, “A novel hybrid technique for acoustic echo cancellation and noise reduction using lms filter and anfis based nonlin- ear filter,” 2016.

[3] S. K. Dubey and N. K. Rout, “Flms algorithm for acoustic echo cancellation and its comparison with lms,” in *2012 1st International Conference on Recent Advances in Information Technology (RAIT)*. IEEE, 2012, pp. 852–856.

[4] K. Kinoshita and T. Nakatani, “Speech dereverberation using linear prediction,” *NTT Technical Review*, vol. 9, no. 7, pp. 1–7, 2011.

[5] M. Parchami, W.-P. Zhu, and B. Champagne, “Speech dereverberation using linear prediction with estimation of early speech spectral variance,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 504–508.

[6] M. Togami and Y. Kawaguchi, “Noise robust speech dereverberation with kalman smoother,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 7447–7451.

[7] U. B. A. Godse, *Basic Electronics Engineering*. Technical Publications, 2008.

[8] A. Yu and C. Mead, “Characteristics of aluminum-silicon schottky barrier diode,” *Solid-State Electronics*, vol. 13, no. 2, pp. 97–104, 1970.

[9] J. M. Fiore, *Operational amplifiers and linear integrated circuits: Theory and application*. West Publishing Company, 1992.

[10] D. A. Bell, *Operational amplifiers and linear ICs*. DA Bell, 1997.

[11] C. Noviello, “Mastering stm32,” *Leadpub. Obtenido de http://www2. keil. com/mdk5/uvision*, 2017.