Noise Suppression In Automatic Speech Recognition Using Hidden Markov Model

Saimir Tola[1)]
Department of Mathematical Engineering
Polytechnic University of Tirana, Albania
E-mail:saimir_tola@yahoo.com

Ligor Nikolla[2)]
Department of Mathematical Engineering
Polytechnic University of Tirana, Albania

***Abstract*—In this paper we propose an Automatic Speech Recognition (ASR) in order to suppress various noises. We try to increase the accuracy to get clear and robust speech recognition. We also are going to use the Discrete Hidden Markov Model (HMM) in order to increase the speed of the recognition. The method of cepstral coefficients are used to increase the robustness of speech recognition. An in-build function in LabView is going to be used to achieve.**

Keywords—Automatic Speech Recognition, cepstral coefficients, Hidden Markov Model

Introduction

Being able to speak to your personal computer, and have it recognize and understand what you say, would provide a comfortable and natural form of communication. It would reduce the amount of typing you have to do, leave your hands free, and allow you to move away from the terminal or screen. You would not even have to be in the line of sight of the terminal. It would also help in some cases if the computer could tell who was speaking. Limited forms of speech recognition are available on personal workstations. Currently there is much interest in speech recognition, and performance is improving. Speech recognition has already proven useful for certain applications, such as telephone voice-response systems for selecting services or information, digit recognition for cellular phones, and data entry while walking around a railway yard or clambering over a jet engine during an inspection. The improvement of the recognition accuracy robustness may affect the performance of ASR. Mel

Frequency cepstral (MFCC), Perceptual linear prediction (PLP), Linear Predict Coding (LPC) are the various methods of feature extraction techniques available for ASR.

MFCC and PLP are the methods used in order to reconstruct the original signal [1]. We suppress the noise from the signal in order to increase the robustness of the speech recognition.

We sample the speech signal at 16000 Hz and the number of bits per sample is 32.

The modeling of the sample can be done according to the assumption that a small segment of speech signal is sufficiently stationary [2]. The feature extraction technique is shown in figure 1 below :

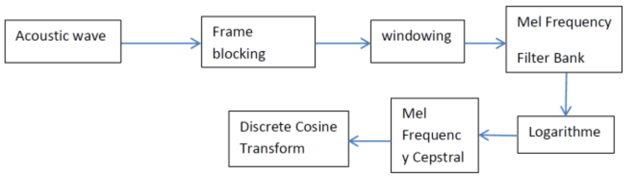

Figure 1: Feature Extraction Technique

We apply frame blocking to divide the signals into Matrix form. The number of frames could be 320 samples within a frame. Over lapping of frames would have the factor of separating of samples due to the effect of frame blocking [3].

After this we apply the hamming windowing in order to reduce the discontinuities of signal at the end of each frame.

The human ear is very sensitive and we know that humans have high resolution to the low frequency.

And we use Discrete Hidden Markov Model in order to accelerate the speed of speech recognition. This paper utilizes this technique as a comparator based on the probability basis, and it improves the efficiency of the desired systems [4].

Results and discussions

The simulation results can be obtained with the help of a Modelsim software. In figure 2 and 3 we see that the authorized speech due to the match-out signal is high, when the feature vectors are similar. Figure 4 shows that the power dissipated from the desired system. The power dissipated for the desired system is 188.18 mw.

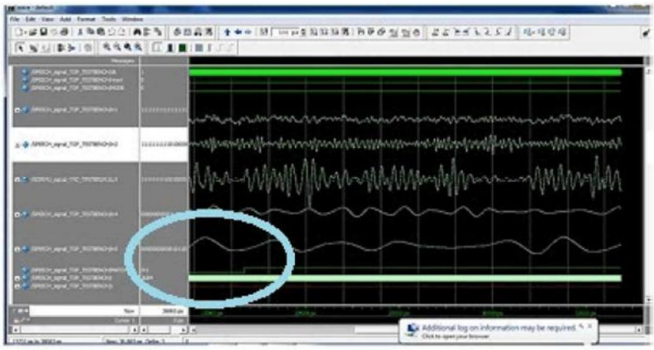

Figure 2: Authorized speech

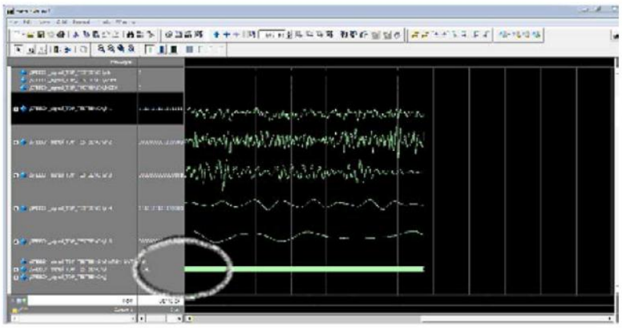

Figure 3: Unauthorized speech

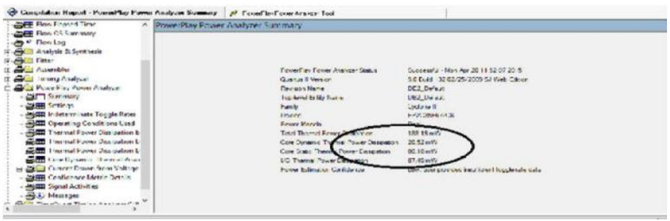

Figure 4 : Power Dissipation

Conclusion

We compare the speech signal with the feature vectors that are obtained by processing the speech. This technique is a slight time consuming process but it provides accuracy for robust speech recognition.

References

[1] Yuan Mang, 2004. Speech Recognition on DSP: Algorithm on optimization & performance analysis, The Chinese University of Hong Kong, pp:1-18.

[2] Huggins-Daines D., M. Kumar, A. Chan, A. Black, M. Ravishekar and A. Rudnicky, 2006. Pocketsphinx: A free, real time continuous speech recognition system for hand-held devices, in Proceedings of ICASSP.

[3] Joshi, Siddhant, C. and Dr. A.N. Cheeran, 2014. MATLAB Based Feature Extraction using Mel Frequency Cepstrum coefficients for automatic speech recognition, IJSETR, 3(6).

[4] Pan Shing-Tai and Xu-Yu Li, 2012. An FPGA Based Embedded Robust Speech Recognition System Designed By Combining Empirical Mode Decomposition and a Genetic Algorithm, IEEE Trans on Instrumentation and Measurement, 61(9).