

Digital Signal Processing

**Performance Analysis and Development of a simple Speech Recognition program in MATLAB**

Project Proposal

Done by

Zubier Abdullah

ID – 500419866

Prepared for

Dr. Karthi Umapathy

EE8105

**Background**

A widely used application in Digital Signal Processing is that of speech recognition. This is because it enables users to access their devices in a much more convenient way. To many people, especially those who are not computer literate, using a computer or a smartphone is difficult and confusing and so having a voice based system, capable of recognizing words and commands as input, makes the process of using the device much simpler and accessible. In recent years, with the improvement of computer technology, companies such as Amazon, Google and Apple are able to provide extremely intelligent and robust voice activated applications to enhance and improve the lives of their consumers.

**Objective**

The objectives of this project are two fold

* Develop a speech recognition in MATAB capable of distinguishing words from different speakers
* Evaluate the noise performance of the system with additive white Gaussian noise being added to the input signal with noise power 0dB, 3dB, 5dBand 6 dB.

**Methods**

This speech identification algorithm will be based on cross correlation and comparison of the reference signal power density spectrum with the input signal power density spectrum. After cross correlation of the spectrums are completed, the system will evaluate how closely the words match each other according to the symmetry of the cross correlation function and identify the word spoken.

The database of words will be taken from audio signals from the internet and will comprise of five reference words. The input word will be compared to the five reference words and the system will identify what word has been spoken. This test will also be replicated with added noise given to the input signals. The performance of the system will be evaluated based on how accurately the system can identify whether two words match or not.

For clarity regarding the methodology used, consider the following table, which will have 5 reference words.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Reference word /Input word | Noise level 0dB | | Noise level 3dB | | Noise level 5dB | | Noise level 6dB | |
| Same Speaker | Different Speaker | Same Speaker | Different Speaker | Same Speaker | Different Speaker | Same Speaker | Different Speaker |
| Cat |  |  |  |  |  |  |  |  |
| Dog |  |  |  |  |  |  |  |  |
| House |  |  |  |  |  |  |  |  |
| Hello |  |  |  |  |  |  |  |  |
| World |  |  |  |  |  |  |  |  |

The algorithm will essentially compute the spectrum of the reference signal and the input signal. This will be done using the spectrogram function. The spectrogram function will be transposed and summed to get a row vector containing the column su mmation results. This row vector is the frequency spectrum signal. The spectrum will be linearly normalized in Matlab so that the range of values will lie within the range [0,1].

The cross correlations between the reference word and the input word will be computed then. We know that when doing cross correlation with two signals, if there is no time shift between the signals, there will be a peak in the middle of the cross correlated signal. However, if there is time shift between them, the cross correlated signal will find the peak elsewhere. We know that the cross correlated signal will be laterally symmetrical about the peak. We firstly compare the time shift between the input word and the reference signals. The closer the time shift between the two words, the more similar the signals are. Finally, we use the symmetry property of the signals and compute the mean squared error between the left side of the spectrum and the right side of the spectrum about the peak value. The system will then use this to identify what is the word which has been spoken.

The database size will consist of five reference words and the input words will also be the reference words, spoken by the same speaker and different speakers. The system performance will be gauged by how accurately it can determine what word has been used as the input

**Expected Outcomes**

The system will be able to identify the input word by comparison with the 5 reference words. So if we compare each word with itself and the other reference words, from the same speaker and from a different speaker, there will be forty comparisons in total.

It is expected that the system will be most accurate when the input word and reference word are spoken by the same speaker. It is also expected that the accuracy will deteriorate as more and more noise is added to the input signals when comparing with the reference signals.

As mentioned before, the system will gauge what word has been spoken by comparing the mean squared error of the cross correlation of the frequency spectrums. The quantifiable outcome of the system is how accurately it can identify the word at different noise levels.

**References**

1. D.Ning, *“Developing an Isolated Word Recognition System in MATLAB.”* MATLAB Digest. Retrieved from www.mathworks.com

2. A. Davletcharova, S. Sugathan, B. Abraham and A.P. James, *“Detection and Analysis of Emotion from Speech Signals”,* Procedia Computer Science (58)

3. Yang T. (2012), *“The Algorithms of Speech Recognition, Programming and Simulating in MATLAB”* (Bachelor’s Thesis), University of Gavle, Gavle, Sweden.

4. Adarsh K.P, A.R. Deepak, Diwakar R. and Karthik R. (2007), *“Implementation of a voice based Biometric System”*, Visveswaraya Technological University, Belagum, Bangalore, India